DOCUMENT RESUME

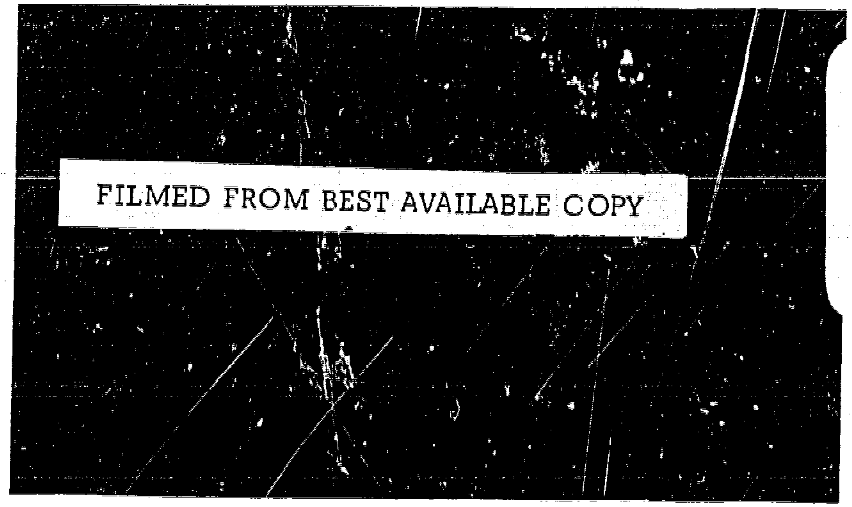
ED 073 942 SE 015 849

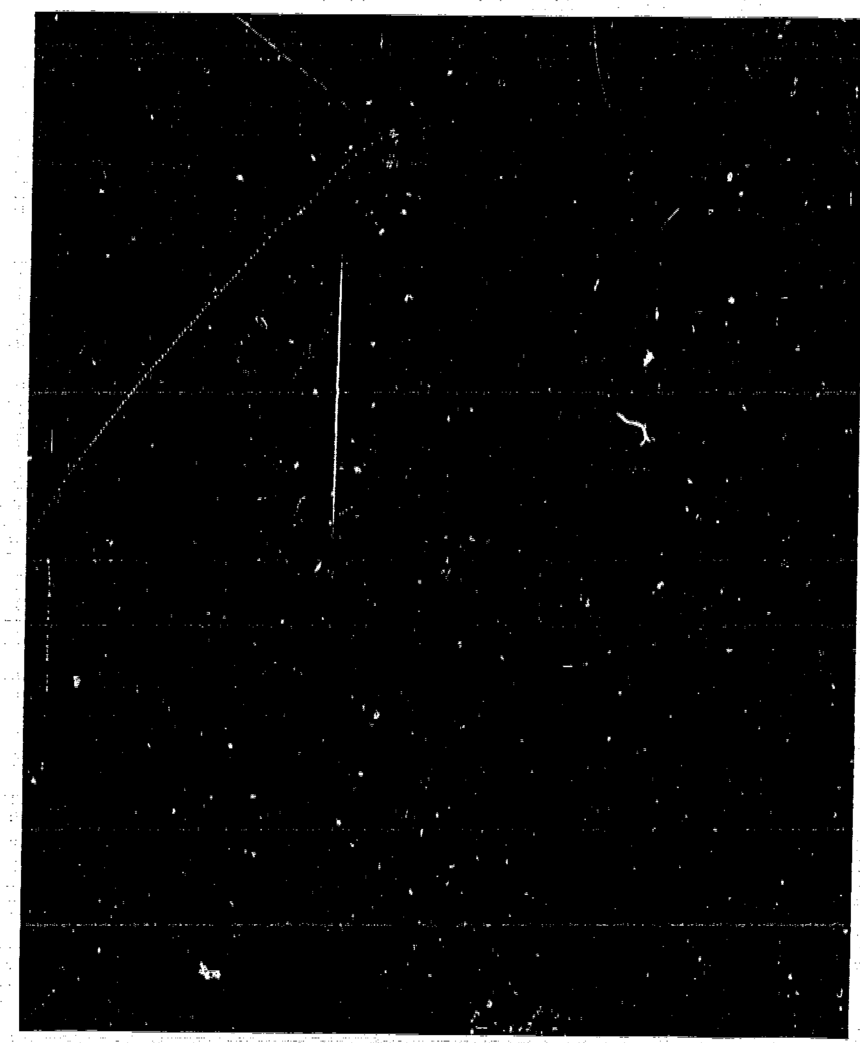
AUTHCR Feldman, Katherine Vorwerk
TITLE The Effects of Number of Positive and Negative Instances, Concept Definition, and Emphasis of Relevant Attributes in the Attainment of Mathematical Concepts.
INSTITUTION Wisconsin Univ., Madison. Research and Development Center for Cognitive Learning.
SPONS AGENCY Office of Education (DHEW), Washington, D.C. Research and Development Centers Branch.
REPORT NO TR-243
BUREAU NC BR-5-0216
PUB CATE Nov 72
CONTRACT OEC-5-10-154
NOTE 120p.

EDRS PRICE MF-$0.65 HC-$6.58
DESCRIPTORS Cognitive Processes; *Concept Formation; Elementary School Mathematics; Geometric Concepts; *Learning; Learning Theories; *Mathematics Education; *Research

ABSTRACT

Three experiments were conducted to determine the effect of several instructional variables on concept attainment. In Experiment I, the effect of presenting a rationally chosen set of positive and negative instances was contrasted with the effect of presenting either a rationally chosen set of positive instances alone or two randomly selected positive instances. Experiment II modified the treatment by giving a concept definition in addition to teaching instances. In Experiment III the effect of presenting a rationally chosen set of positive and negative instances alone was compared with the effect of also including either a concept definition alone or a concept definition plus emphasis of relevant attributes. About 100 sixth grade students took a series of printed lessons dealing with geometric symmetry. Results showed that providing a rational set of positive and negative instances with a definition, or with a definition plus emphasis, was significantly more facilitative in promoting concept learning than the rational set alone, and that providing a rational set of instances with a definition and emphasis was not sifnificantly more facilitative than the rational set with definition alone. (Author/DT)





U.S. Office of Education Center No. C-03 Contract OE 5-10-154

Technical Report No. 243

THE EFFECTS OF NUMBER OF POSITIVE AND NEGATIVE INSTANCES, CONCEPT DEFINITION, AND EMPHASIS OF RELEVANT ATTRIBUTES IN THE ATTAINMENT OF MATHEMATICAL CONCEPTS

Report from the Conditions of Learning and Instruction Component of Program 1

by Katherine Vorwerk Feldman

Vernon L. Allen, Frank H. Farley,
Herbert J. Klausmeier, and Joel R. Levin
Principal Investigators

Wisconsin Research and Development
Center for Cognitive Learning
The University of Wisconsin
Madison, Wisconsin

November 1972

This Technical Report is a master's thesis reporting research supported by the Wisconsin Research and Development Center for Cognitive Learning. Since it has been approved by a University Examining Committee, it has not been reviewed by the Center. It is published by the Center as a record of some of the Center's activities and as a service to the student. The bound original is in The University of Wisconsin Memorial Library.

Published by the Wisconsin Research and Development Center for Cognitive Learning, supported in part as a research and development center by funds from the United States Office of Education, Department of Health, Education, and Welfare. The opinions expressed herein do not necessarily reflect the position or policy of the Office of Education and no official endorsement by the Office of Education should be inferred.

Center No. C-03 / Contract OE 5-10-154

STATEMENT OF FOCUS

Individually Guided Education (IGE) is a new comprehensive system of elementary education. The following components of the IGE system are in varying stages of development and implementation: a new organization for instruction and related administrative arrangements; a model of instructional programing for the individual student; and curriculum components in prereading, reading, mathematics, motivation, and environmental education. The development of other curriculum components, of a system for managing instruction by computer, and of instructional strategies is needed to complete the system. Continuing programmatic research is required to provide a sound knowledge base for the components under development and for improved second generation components. Finally, systematic implementation is essential so that the products will function properly in the IGE schools.

The Center plans and carries out the research, development, and implementation components of its IGE program in this sequence: (1) identify the needs and delimit the component problem area; (2) assess the possible constraints--financial resources and availability of staff; (3) formulate general plans and specific procedures for solving the problems; (4) secure and allocate human and material resources to carry out the plans; (5) provide for effective communication among personnel and efficient management of activities and resources; and (6) evaluate the effectiveness of each activity and its contribution to the total program and correct any difficulties through feedback mechanisms and appropriate management techniques.

A self-renewing system of elementary education is projected in each participating elementary school, i.e., one which is less dependent on external sources for direction and is more responsive to the needs of the children attending each particular school. In the IGE schools, Center-developed and other curriculum products compatible with the Center's instructional programing model will lead to higher morale and job satisfaction among educational personnel. Each developmental product makes its unique contribution to IGE as it is implemented in the schools. The various research components add to the knowledge of Center practitioners, developers, and theorists.

TABLE OF CONTENTS

Page

LIST OF TABLES

LIST OF TABLES (continued)

ABSTRACT

Three experiments were conducted to determine the effect of several instructional variables on concept attainment. Experiment I focused on the role of positive and negative teaching instances. The effect of presenting a rationally chosen set of positive and negative instances was contrasted with the effect of presenting either a rationally chosen set of positive instances alone or two randomly selected positive instances. A control group read placebo lessons. Experiment II was a modification of Experiment I; in each treatment condition a concept definition was given in addition to teaching instances. In Experiment III the effect of presenting a rationally chosen set of positive and negative instances alone was compared with the effect of presenting this rational set of instances with (1) a concept definition and (2) a concept definition and emphasis of relevant attributes. A control group again read placebo lessons.

Approximately 100 sixth-grade students participated in each experiment as subjects. The instructional variables were manipulated in a series of printed lessons dealing with geometric symmetry.

Although five dependent variables were used to assess both immediate concept acquisition and retention, only the subjects' ability to correctly identify new instances proved to be a consistently reliable measure. Results for this variable showed that:

(1) providing a rational set of positive and negative teaching instances resulted in significantly better performance than that of the control group (immediate acquisition only).

(2) providing only a rational set of positive instances or two randomly chosen positive instances did not result in significantly better performance than that of the control group (immediate acquisition and retention).

(3) presenting a concept definition in addition to teaching instances (rational set of both positive and negative instances, rational set of positive instances only, or two positive instances) resulted in equal performance among treatment conditions and significantly better performance by all treatment groups than that of the control group (immediate acquisition and retention).

(4) providing a rational set of positive and negative instances with a definition plus emphasis resulted in significantly better performance than that of the control group (immediate acquisition and retention).

(5) providing a rational set of positive and negative instances with a definition, or with a definition plus emphasis, was significantly more facilitative in promoting concept learning than the rational set alone (immediate acquisition only).

(6) providing a rational set of instances with a definition and emphasis was not significantly more facilitative than the rational set with a definition alone (immediate acquisition and retention).

Chapter I

INTRODUCTION

In recent years a great many studies dealing with the nature of concept learning have been conducted. Impetus for this research has come both from learning theorists, who study concept learning within a purely theoretical framework, and from educators, who recognize concept attainment as a fundamental form of classroom learning.

The majority of the research on concept learning has been conducted in the laboratory under highly controlled conditions. These laboratory studies have made significant contributions to the psychological theory of concept learning as well as contributing to the understanding of how concepts are learned outside of the laboratory. However, the laboratory research has been limited in its scope. Typically only concepts with highly perceptible attributes have been studied, and the instructional techniques used have generally not extended beyond providing subjects with concept examples and non-examples. Furthermore, only a limited number of behaviors, such as the ability to correctly identify a criterion number of concept examples, have been used to infer concept attainment.

Clearly, much of the concept learning which occurs outside of the psychological laboratory does not fit into the limited framework of the typical laboratory study. In the classroom, for example, concepts of varying degrees of abstractness are introduced, and they are not taught merely by providing examples and non-examples, but also through the use of labels or names, definitions and synonyms. Additionally, attainment of the concepts is inferred from a variety of behaviors, such as the ability to correctly define the concept and recognize the relationships between it and other concepts.

Due to the limited scope of the majority of the studies on concept attainment, researchers working at the Wisconsin Research and Development Center for Cognitive Learning have undertaken a comprehensive program of research designed to increase knowledge about the stimulus variables and cognitive operations related to concept attainment both in the laboratory and in the classroom situation. One of the products of this research has been the formulation by Klausmeier (1971) of a descriptive model of the cognitive operations involved in concept learning. This model views concept learning as a complex form of learning which can be subdivided into four hierarchical levels, each level representing knowledge about concepts at a higher degree of inclusiveness and abstractness.

The four levels of concept learning postulated in Klausmeier's model are concrete, identity, classificatory, and formal. At the concrete level the individual is able to recognize an object which he has experienced earlier. At the identity level the individual recognizes the object even when viewed from a different perspective or sensed in a different modality. At the classificatory level the individual can identify at least two different instances of a concept as belonging to the same set or class even though he cannot name the attributes common to them. Finally, at the formal level the individual can identify examples and non-examples of the concept, name the concept, and identify it in terms of its relevant attributes.

Researchers at the Center have also identified a taxonomy of the variables which they believe influence concept learning (Klausmeier, Davis, Ramsay, Fredrick, & Davies, 1965). These variables are hypothesized to affect learning at each of the four levels in Klausmeier's model. The taxonomy has three major classifications: instructional variables, learner variables, and concept variables. Instructional variables are those related to the manner in which the concept is presented, such as the presence or absence of a definition or the sequence of presenting examples and non-examples. The learner variables classification refers to variables characteristic of the subject, such as age or IQ, and includes the various cognitive operations that any particular individual may or may not be able to carry out. Concept variables are those relating to the nature of the specific concept being learned, including its relevant and irrelevant attributes, examples and non-examples.

Currently a variety ... dies are underway to establish both the validity of the concept learning model and the specific effect on concept attainment of the variables in the taxonomy at each of the four levels indicated in the model. The present study is one of these research efforts.

Purposes and Hypotheses of the Study

The purpose of the present study was to focus on concept learning at the formal level by manipulating within printed materials three of the instructional variables outlined in the taxonomy and measuring their effect on several of the behaviors from which attainment at the formal level can be inferred. The instructional variables were:

(1) The number of examples and non-examples of the concept presented when the concept is not defined.

(2) The number of examples and non-examples of the concept presented when the concept is defined.

(3) Emphasis of the relevant attribute values of the concept.

A variety of hypotheses were tested and they are presented in detail in Chapter IV. Very broadly, however, it was hypothesized that (1) providing a "maximal and optimal"* number of concept examples and non-examples would be more facilitative than

---

*"Maximal and optimal" refers to what Markle and Tiemann (1969) have called the rational set of teaching examples and non-examples. This concept is defined in full in Chapter II.

a lesser number only if the concept were not defined, and (2) that emphasis of the relevant attribute values of the concept would facilitate concept learning.

Method

The instructional variables were studied separately in three independent experiments. In each experiment three selected geometric concepts were presented through printed lessons. The same concepts were used in all three studies. The lessons prepared for each experiment were designed to systematically vary the particular instructional variable focused on in the experiment.

The three experiments used an identical series of tests. The tests consisted of a variety of questions related to concept learning at the formal level.

The subjects were sixth-grade students. Ninety-six subjects participated in Experiment I, 118 in Experiment II, and 108 in Experiment III.

The procedure followed in each of the experiments was essentially identical. The experimental materials were administered in two sessions. During the first session subjects read through the printed lessons and took a series of tests based on the lessons. Two weeks later the subjects were tested in the second session for the retention of the concepts.

Significance of the Study

Knowledge of the effect on concept learning at the formal level of each of the three instructional variables studied in the present series of experiments could have important implications for the theory of concept learning, particularly as advanced by Klausmeier's model. Moreover, by using several dependent measures to infer concept attainment it will be possible to determine the effect of the instructional variables on each of these several aspects of concept learning.

The present study also has potential significance for educators and textbook writers concerned with presenting concepts effectively through printed materials. The results of this study and other similar studies dealing with the effect of within-text instructional variables on concept attainment could form the basis for a set of guidelines for effective concept presentation. Such a set of prescriptions would be a valuable advance in the preparation of classroom materials, as the current methods used are based entirely on anecdotal evidence, untested theories, or laboratory studies which are not directly applicable to the teaching situation.

Chapter II

REVIEW OF RELATED LITERATURE

Teaching Concepts

Very little work, either theoretical or empirical, has focused on the problem of how to teach concepts effectively. The only practical although partially developed theories in the area are those advanced by Markle and Tiemann (1969) and Merrill (1971). As these theoretical frameworks are directly related to the present study, they will be discussed in some detail.

Markle and Tiemann (1969) postulate that teaching a concept actually involves teaching two independent behaviors. These behaviors are: (1) generalizing among instances of the concept class, and (2) discriminating instances of the concept class from instances of other classes. The teaching of each of these behaviors according to Markle and Tiemann necessitates the use of different instructional variables. To teach a student to generalize within a class or concept the student must be presented with enough examples (positive instances) of the concept to vary each major irrelevant attribute of the concept. Irrelevant attributes of a concept are those attributes which may or may not be common to every example of the concept and which are not essential to defining the concept such as color for the concept "balloon." To teach a student to discriminate between concepts, however, the student must be provided with enough non-examples (negative instances) of the concept to systematically exclude each relevant attribute. Relevant attributes of a concept are the properties which are common to every example of the concept and, therefore, are the properties which define the concept such as one pair of parallel sides for the concept "trapezoid." The number of examples and non-examples needed to insure both generalization and discrimination, which is of course dependent upon the number of relevant and irrelevant attributes of the specific concept being taught, is called the "rational set of teaching examples and non-examples." If the student is presented with the rational set of both examples and non-examples, Markle and Tiemann theorize that he will be able to correctly identify other concept examples and non-examples.

When the rational set of teaching examples and non-examples is not provided, Markle and Tiemann predict that the subject will make certain classification errors in identifying new examples and non-examples. These errors are:

1. Overgeneralization: The student identifies some non-examples of the concept as examples. Overgeneralization is the result of providing the student with an insufficient variety of non-examples.

2. Undergeneralization: The student identifies some examples of the concept as non-examples. Undergeneralization is the result of providing the student with an insufficient variety of examples.

3. Misconception: The student identifies some non-examples as examples and some examples as non-examples. Misconception is the result of providing the student with an insufficient number of

both examples and non-examples. The student is classifying on the basis of an irrelevant attribute.

Unfortunately, no research has been undertaken to specifically validate Markle and Tiemann's hypotheses concerning the effectiveness of the rational set in promoting concept learning and the occurrence of the specified classification errors which result if this optimal number of teaching instances is not provided. One of the purposes of the present study was to provide empirical data on these hypotheses by contrasting the effects on concept learning of providing the rational set of both examples and non-examples, the rational set of examples, and just two examples.

Merrill (1971) has proposed a method of teaching concepts which is similar to that of Markle and Tiemann. Merrill concurs that teaching examples of the concept which differ widely in irrelevant attributes are necessary to promote generalization. However, he postulates that discrimination is promoted by providing teaching non-examples of the concept which have irrevelant attributes identical to those of the teaching examples.

Some evidence for the validity of the theoretical work of both Markle and Tiemann and Merrill comes from the results of a study by Tennyson, Woolley, and Merrill (1972). Thirty-five college students were presented with a definition of the concept <u>trochaic meter</u> followed by 16 labeled examples and non-examples of the concept. Concept attainment was measured on an acquisition test requiring the subjects to identify 30 new examples and non-examples.

Three independent variables were manipulated in the experiment: divergency, matching, and probability. The Divergency variable was developed from Merrill's (and Markle and Tiemann's) hypothesis concerning the type of positive instances which promote generalization. A divergent condition was defined as one in which the irrelevant attributes of the positive instances were varied as much as possible. Similarly, matching was developed from Merrill's theory of the kind of non-examples which promote discrimination. A matched condition was one in which the teaching examples and non-examples had similar irrelevant attributes. Probability referred to the difficulty of the teaching examples and non-examples. Probability "scores" had been assigned to each instance on the basis of an "instance probability analysis" in which an earlier and independent group of subjects had identified the instances as examples or non-examples of trochaic meter based only on a concept definition. The probability rating of each instance was defined as the percentage of subjects who had been able to correctly identify it from the defintion. Subjects presented with only high probability or obvious teaching instances were expected to identify only obvious examples on the acquisition test thus making many undergeneralization errors. Conversely, subjects presented with only low probability items were expected to overgeneralize by identifying almost every item on the test as an example.

Through logical manipulations of the three independent variables, Tennyson *et al*. generated four treatment conditions. For each condition a particular outcome was hypothesized. These outcomes were either correct classification (all positive and negative instances correctly identified) or one of the three classification errors postulated by Markle and Tiemann. The treatment conditions for the experiment were:

1. High to low probability instances, divergent, and matched. Hypothesized outcome: correct classification.

2. Low probability instances, divergent, and not matched. Hypothesized outcome: overgeneralization.

3. High probability, divergent, and matched. Hypothesized outcome: undergeneralization.

4. High to low probability, convergent, and not matched. Hypothesized outcome: misconception.

The results of the experiment supported every hypothesis ($p<.01$). Of particular interest is a comparison of the outcomes of Conditions 1 and 4. When the teaching examples and non-examples were selected according to rational criteria (Condition 1), the subjects were able to correctly classify new instances on the acquisition test; however, when the examples and non-examples were in effect randomly selected (Condition 4), the subjects were not able to correctly classify new instances and made both over- and undergeneralization classification errors. Clearly selecting teaching instances on the basis of rational criteria, such as those advanced by Merrill and Markle and Tiemann, is potentially an important instructional variable in concept learning.

The extant research on concept learning, exclusive of that conducted by Merrill and his associates, does not unequivocably support the conclusions reached by Markle and Tiemann and Merrill on the specific use and effectiveness of examples and non-examples in concept learning. These differences, along with possible explanations, will be discussed in the following sections in which the concept learning research dealing with the three independent variables focused on in the present study will be reviewed.

Number of Positive Instances

The role of positive instances in concept learning according to both Markle and Tiemann and Merrill is to promote generalization. This is accomplished by presenting the subject with enough teaching examples to sufficiently vary the major irrelevant attributes of the concept. Therefore, the number of examples needed is directly dependent upon the concept in question. However, the concept learning research dealing with the "optimal" number of positive teaching instances has focused primarily on an absolute number unrelated to any particular concept. The results of these studies have been inconclusive as to whether a large number or a small number of examples is more facilitative in concept learning.

Podell (1958) presented college students with either two or twelve examples of a figural concept and instructed them to discern their common attributes. Later when asked to list the common features of the designs, the subjects who had seen twelve examples were able to recall significantly more relevant attributes than subjects who had seen only two examples.

A study by Amster and Marascuilo (1965) yielded results nearly contradictory to those reported by Podell. Fourth-grade children were taught the concepts of set-union and set-interaction using either twelve or thirty-six different examples. No difference was found in the number of correct responses made on a learning task due to the number of different examples presented, although subjects who had learned the concepts from a small number of examples performed significantly better on a generalization task which employed words or letters as instances of the concepts.

Stern (1965) found that training with an intermediate number of examples was most beneficial. Kindergarten and first-grade children were presented with either three instances of each of eight concepts, six instances of each of four concepts, or two instances of twelve concepts. The intermediate number of instances (six instances of four concepts) facilitated transfer to both new instances and new concepts for the older subjects.

Frayer (1970) found no differences at all due to the number of examples provided. She presented geometric concepts to fourth and sixth-grade children through programed lessons using either two or four positive instances. After studying the lessons, the subjects were administered a test consisting of eleven types of questions related to concept learning. The level of concept mastery of subjects who had learned the concepts from two examples was not significantly different from that of the subjects who had learned the concept from four.

It is likely that the differences among the results of the studies dealing with the "optimal" number of positive instances are due to the fact that the number of examples necessary for concept learning is not an absolute number or merely a matter of being large or small. Rather, it is probable that this number is dependent on the specific concept involved as both Markle and Tiemann and Merrill have suggested. This issue will be examined in part in the present series of experiments by comparing the effectiveness of only two positive instances with the rational set of positive instances.

Number of Negative Instances

The literature on negative instances has not been directly concerned with the "optimal" number of negative instances necessary to ensure concept attainment, but rather with whether non-examples are facilitative at all in promoting learning. Indeed, a great deal of research has been undertaken to answer this question, and, overall, this research has not shown non-examples to be facilitative.

The earliest study dealing with negative instances was reported by Smoke (1933) who used college students as subjects.

He contrasted the performance of subjects who learned a concept from a series of labeled positive instances with the performance of subjects who learned the concept from a series of equal numbers of labeled positive and negative instances. Smoke found no statistical differences between the two groups on a number of dependent measures and concluded that negative instances do not, at least, retard learning.

Hovland (1952) criticized Smoke's study by pointing out that the amount of information about a concept conveyed in a non-example is far less than that conveyed in an example. He argued that the information content of the two types of instances should be equated in concept learning problems and developed a method to do this. Subsequently, Hovland and Weiss (1953) conducted a series of experiments in which the information content of the examples and non-examples was equated using Hovland's technique. The results showed that significantly more subjects were able to solve problems defined by all positive instances than those defined by all negative instances. The percentage of subjects solving problems defined by equal ratios of positive to negative instances was at an intermediate level. Because the information content of the instances was equated, Hovland and Weiss concluded that the differences in concept attainment were due to difficulties in using the information conveyed by negative instances.

In a recent study by Smuckler (1967), the results of the Hovland and Weiss experiments were confirmed, although no attempt to equate the information content of the instances was made. The subjects, 80 second graders, learned the concept "trapezoid" from 40 labeled teaching instances. At eight-figure intervals (trials) throughout the series of 40 instances the subjects were tested on unlabeled figures. Four ratios of positive to negative instances were used in the 40 teaching instances: 100, 75, 50, and 25 per cent positive instances. The 100% positive instances condition resulted in a consistently greater percentage of correct responses on intertrial identifications as well as on a transfer task requiring the subjects to identify 30 new instances of "trapezoid."

Bourne (1966) has argued that a distinction must be made between concepts defined by a conjunctive rule and concepts defined by a disjunctive rule when evaluating the comparative utility of using positive and negative instances. Bourne points out that most of the studies dealing with negative instances have involved conjunctive concepts, that is concepts with at least two relevant attributes both of which must be present in every example. When the concept is conjunctive, the examples simply carry more logical information about the concept then do the non-examples because each example defines precisely what the concept is. Therefore, their information is more easily assimilated. However, when the concept is disjunctive, so that an instance is an example of the concept if it contains at least one of the relevant concept attributes, the information value of the examples and non-examples is reversed. The negative instances contain more logical concept information because they precisely define what the concept is not, whereas the positive instances merely define one aspect of what the concept is.

In a recent survey of the research on concept learning, Clark (1971) has shown that the literature supports Bourne's conclusions about the utility of negative instances. Only five out of 25 studies reviewed found that a sequence of positive instances or a sequence of positive and negative instances was more effective than a series of all positive instances in promoting learning when the concept was conjunctive, whereas four out of four studies dealing with disjunctive concepts concluded that a series of all negative instances increased concept attainment when compared to a sequence of all positive instances.

A study by Huttlenlocher (1962), however, again complicates the issue of negative instances. Huttenlocher contrasted the effects on concept learning of a series of all positive instances, all negative instances, or an equal number of both positive and negative instances when the concept involved had only one relevant attribute, therefore neither conjunctive nor disjunctive. Contrary to other research, performance with a mixed series (non-examples followed by examples) was found to be superior.

Tennyson (1971) suggests that the traditional laboratory studies have not found negative instances to be uniformly facilitative because these studies have not presented negative instances properly. He argues that negative instances only facilitate concept learning when they are used in a way which forces the subjects to concentrate on the relevant attributes of the concept. It is this function which promotes discrimination, and it is this which the previous research has failed to do. Tennyson points out that one way to use negative instances to focus the subjects attention on the relevant attributes is to present examples and non-examples in a matched relationship (matched on irrelevant attributes), the technique used in the Tennyson, Woolley, and Merrill study. By matching the instances in this way the subjects are forced to concentrate on the relevant concept attributes because they are the only differences between the examples and the non-examples.

To point out the facilitative effects of the negative instances in the Tennyson, Woolley, and Merrill study, Tennyson (1971) essentially replicated the Tennyson et al. experiment but eliminated the matching variable by using only positive instances. The results showed that without negative instances the subjects (seventh-grade students) responsed randomly on the post test requiring them to identify new instances of the selected concept (adverb).

The results of Tennyson's study also lent support to Markle and Tiemann's position concerning the use of negative instances in concept learning. For their suggestion that negative instances be used to vary the relevant concept attributes does, like the Tennyson et al. matching variable, force the subject to concentrate on the relevant attributes of the concept as exemplified in the positive instances. The utility of non-examples in the Markle and Tiemann paradigm was specifically studied in the present experiment by comparing the effects of presenting the rational set of both positive and negative instances with the effects of presenting only the rational set of positive instances.

Concept Definition

The definition of a concept can be looked upon as a list of the relevant attributes of the concept and the rule by which they are combined. Thus the definition conveys the same information as an example of a conjunctive concept; it states precisely what the concept is. But the definition also carries additional information in that by omission it states what the concept is not. It is reasonable to assume, therefore, that presenting a concept definition in terms of relevant attributes and the rule by which they are combined is potentially a powerful instructional variable in effecting concept attainment.

In a study by Anderson and Kulhavy (1972) it was shown that a high degree of concept attainment results from merely presenting the concept definition. Subjects (college students) were provided with one-sentence definitions of unfamiliar concepts and instructed to use the definitions to correctly identify examples of these concepts on a multiple choice test. Subjects in a control condition also took the test but were not provided with the concept definitions. The error rate of the group without definitions was 71.3%, while that of the group with definitions was 7.0%.

Merrill and Tennyson (1971) in a study employing the same experimental paradigm used in the Tennyson *et al.* experiment found that concept definitions facilitate learning even when labeled positive and negative instances are also given. Subjects provided with teaching examples and non-examples and a concept definition performed significantly better than subjects provided only with examples and non-examples.

Both the Anderson and Kulhavy study and the Tennyson study indicate that the presence of a concept definition is highly facilitative. The study by Frayer (1970) suggests further that the effect of providing a definition may be powerful enough to eliminate any differential effects due to the number of instances provided. She found no significant differences between subjects provided with a concept definition and four instances (two negative and two positive) and subjects provided with a definition and eight instances (four negative and four positive). However, no other research dealing with the effect of a concept definition when presented with various numbers of teaching examples and non-examples has been reported. One of the purposes of the present study was to investigate this variable. The effect of presenting a definition with the rational set of both positive and negative instances was compared to (1) the effect of presenting a definition with the rational set of positive instances, and (2) the effect of presenting a definition with just two positive instances. It was hypothesized that the concept definition would be so facilitative that varying the number of teaching instances would have no significant effect on concept learning.

Emphasis of Relevant Attributes

The use of verbal cues to draw attention to the relevant attributes of concept examples has repeatedly been shown to facilitate concept learning. Gelfand (1958) had three groups of college students memorize lists of words prior to a concept identification test. One group of subjects memorized words describing the relevant attributes of the concept, another words describing the irrelevant attributes, and a third neutral words unrelated to the concept. Performance on the identification task was significantly better for subjects who had memorized the list of words describing the relevant attributes of the concept.

Wittrock, Keislar, and Stern (1964) investigated the effects on concept identification of providing general, class, or specific cues during a training period. The subjects were kindergarten children who had previously been taught a hierarchy of verbal associations over a three-month period. The word "article" had been associated with the words La and Le, each of which had in turn been associated with six French nouns. The concept identification task involved pictures of these twelve nouns. The subjects' task was to match one of two pictures with a model picture on the basis of the French name for the model. During the task-training period involving pictures of only six of the nouns, the subjects were either (1) told that the "article" was the basis for matching (general cue), (2) given the specific article which was the basis for the matching (class cue), (3) given the French name of both the model and the matching picture (specific cue), or (4) supplied with no cue at all. Subjects who had been given class cues during training performed significantly better on immediate transfer, delayed transfer, and retention tests (involving the remaining six pictures or different combinations of the pictures used in training) than subjects given general cues, specific cues, or no cues.

The presence or absence of a verbal cue drawing attention to relevant concept attributes was also one of the variables investigated by Remstad (1969). He presented examples with either the concept's name or the name and a one word clue referring to a relevant attribute. The presence of single word cues greatly increased performance on a transfer task.

Frayer (1970) used attention directing and review questions to focus the subjects (4th and 6th graders) attention on the relevant concept attributes. She found that emphasis of relevant attributes significantly increased overall concept mastery for fourth graders. Additionally, recognition and production of attribute names for 4th graders and recognition of attribute names for 6th graders increased significantly when relevant attributes were emphasized.

In the Merrill and Tennyson (1971) study, attention was drawn to relevant attributes by identifying the relevant attributes in each example of the concept and the absence of the relevant attributes in each non-example. Merrill and Tennyson found this to be a very powerful variable. The error rate on a transfer task dropped significantly when positive and negative instances were presented with emphasis.

In summary, emphasizing relevant attributes by using attention directing verbal cues has consistently facilitated concept learning.

In the present study the "strength" of this facilitative effect was investigated by determining whether emphasis significantly improves concept attainment when a definition and the rational set of both examples and non-examples are also provided.

Chapter III

DEVELOPMENT OF INSTRUCTIONAL MATERIALS AND DEPENDENT MEASURES

Three geometric concepts were chosen as the subject matter for the present study. They were bilateral symmetry, rotational symmetry, and translational symmetry. These concepts were selected for the following reasons: they have perceptible instances and easily specifiable relevant and irrelevant attributes; they are related concepts; they were considered of appropriate difficulty for sixth-grade students by mathematics curriculum developers working at the Wisconsin Research and Development Center; and a review of the fifth- and sixth-grade mathematics texts currently in use revealed that they were rarely covered in the normal curriculum.

Each of the three concepts was analyzed by specifying for each relevant and irrelevant attributes, a definition, coordinate and supraordinate concepts, and a large variety of examples and non-examples. Care was taken to insure that several rational sets of examples and non-examples as defined by Markle and Tiemann (1969) were included among the positive and negative instances generated for each concept.

Instance Probability Analysis

An instance probability analysis (similar to that described by Tennyson *et al.*, (1972) was undertaken to provide data on the information content or "obviousness" of each of the examples and non-examples generated for the three concepts. The results of the analysis were used to select examples and non-examples for inclusion in the lessons and tests used in the present series of experiments.

Subjects

The subjects were 111 sixth-grade students. They comprised the entire sixth-grade population of a middle school in Lake Mills, Wisconsin.

Materials

The materials consisted of two parallel forms of the same test (instance probability analysis). Approximately one half of the subjects were given each form.

The test was divided into three sections, one dealing with each of the three experimental concepts (bilateral, rotational, and translational symmetry). In each section the definition of the concept was presented, followed by an array of figures which were either examples or non-examples of the concept. Beneath each figure were printed the words "yes" and "no."

Included in each test booklet before the test itself were a numbered list of words used in the test which the experimenter felt might be unfamiliar to the subjects and a sample item dealing with the concept "food." The instructions for the sample item were identical to those used in the actual test.

Procedure

The subjects' task on all three parts of the test and on the sample item was to read the definition of the concept presented and then on the basis of the definition to decide which of the figures following it were examples of the concept and which were not. The subjects were instructed to circle the word "yes" beneath each figure which they thought was an example of the concept and to circle the word "no" beneath each figure which they thought was not an example.

Prior to actual testing the experimenter went over the numbered word list and the sample item with the subjects. In reviewing the word list, the experimenter began by asking, "Do you see the word ----- listed here?" After several subjects indicated that they saw the word the experimenter said, "What number is it?" The experimenter waited until most of the subjects had located the word and then called on one subject to give the number of the word. The experimenter and the subjects then pronounced the word together. The entire procedure was then repeated until each word on the list had been covered.

Results

The probability of each instance being correctly identified was determined by calculating the percentage of subjects who had correctly identified it as either an example or a non-example of the concept. The resulting "probability rating" for each instance was interpreted as a measure of that instance's "obviousness"; that is, how apparent it was to the subjects that the instance was or was not an example of the concept defined.

The analysis resulted in a range of probabilities for both positive and negative instances. The mean probability rating for both types of instances on all three concepts was approximately .70.

The probability ratings were used to select examples and non-examples for the lessons and tests according to the following criteria:

(1) Only instances with probability ratings near the mean were selected for the concept learning lessons. This insured that the "obviousness" of the instances was held constant across lessons and conditions.

(2) Instances representing a range of probabilities were selected for the mastery tests. This was done because it was felt that subjects who had mastered the concepts should be able to identify instances along the entire continuum of "obviousness."

Lesson and Test Construction

Lessons

Based on the concept analyses and the results of the probability analysis, three basic lessons (Appendix A) were constructed to teach the selected geometric concepts. Lesson I dealt with bilateral symmetry, Lesson II with rotational symmetry, and Lesson III with translational symmetry. The basic lesson for each concept consisted of a concept definition, a rational set of both examples and non-examples, and emphasis of the concept's relevant attribute values. Relevant attribute values were emphasized by pointing out their presence in each example and their absence in each non-example. Variations in the set of basic lessons constituted the treatment conditions for each of the three experiments.

Tests

Five tests (Tests I-V, presented in Appendix B) were constructed and used in each of the three experiments. The tests were designed to assess concept learning at the formal level by measuring three of the behaviors from which concept attainment at this level can be inferred. These behaviors were correct classification of previously unencountered ex~mples and non-examples, recognition of concept definitions, and knowledge of relationships among concepts. Errors made in identifying examples and non-examples were also tabulated and classified as either over-generalization or undergeneralization errors. This classification constituted a fourth and fifth dependent variable.

Test I was based on Lesson I, Test II on Lesson II, and Test III on Lesson III. Each of Tests I-III was constructed in two parts. Part I required the subjects to identify 20 new concept examples and non-examples, and Part II was a multiple choice item requiring the subject to select the correct concept definition from among four alternative definitions. Test IV was a comprehensive test covering the material presented in all of the lessons and it also was constructed in two parts. Part I consisted of nine items each of which required the subjects to determine whether the item was an example or a non-example of each of the three concepts presented. This part of Test IV was designed to measure the subjects' knowledge of the relationships among the concepts. Part II was a matching item requiring the subjects to match each of the three concepts with its definition. Test V was designed to be used as a measure of retention and consisted of Part I from each of Tests I-III (identification items) and Part II of Test IV (matching-definition item).

Pilot Study

There were three purposes for conducting the pilot study. The first was to determine the appropriateness of the instructional materials for sixth-grade students. The second was to evaluate the quality and clarity of the lessons, tests, and experimenter's instructions to the subjects. The third was to obtain estimates of the time required to administer the materials. Rather than using all of the variations of the concept learning lessons needed for the main studies, only the set of basic lessons was used in the pilot.

Subjects

The subjects were 25 sixth-grade students. They constituted an entire classroom in an elementary school in Menomonee Falls, Wisconsin. The classroom of students was chosen at random from among five sixth-grade classrooms in the school.

Materials

The materials consisted of basic Lessons I-III and Tests I-IV. Included in the booklet for Lesson I was a list of the difficult words contained in the three lessons. A sample item for Part I of Test IV (relational items) was included in the test booklet to insure that the subjects understood the instructions for this part of the test.

Procedure

The list of difficult words was reviewed in a manner similar to that used in the instance probability analysis. Subjects then read the lessons in order, taking each of Tests I-III immediately after completing the lesson upon which it was based. After all of the lessons and their accompanying tests had been completed, Test IV was administered.

Results

The time required to complete the series of lessons and tests was approximately 45 minutes. Ambiguities and other difficulties with the materials and the experimenter's instructions to the subjects were noted and used as a basis for minor revisions.

Mean error rates were calculated for each test. They were: 17%-Test I; 25%-Test II; 26%-Test III; and 51%-Test IV. On the basis of this data it was concluded that the lessons and tests were suitable for sixth-grade students. Although the error rate for Test IV was high, it was nevertheless decided that the test be used in the main studies with only minor revisions as any attempt to make the test easier would have required radical changes in the item types.

Chapter IV

METHOD

Experiment I

The purpose of Experiment I was to determine the effect of the number of teaching instances presented on concept learning when the concept itself is not defined. There were four experimental conditions:

<u>Condition 1</u>: Concept lessons containing the rational set of both examples and non-examples.

<u>Condition 2</u>: Concept lessons containing the rational set of examples only.

<u>Condition 3</u>: Concept lessons containing two examples.

<u>Condition 4</u>: Control condition: placebo lessons.

The dependent measures used to assess concept learning were: (1) correct classification of examples and non-examples; (2) recognition of concept definitions; (3) knowledge of relationships among concepts; (4) overgeneralization classification errors; and (5) undergeneralization classification errors.

Markle and Tiemann (1969) have postulated that providing the rational set of teaching examples and non-examples will assure concept attainment, at least as measured by the ability to correctly classify concept instances. Moreover, they predict that subjects will make overgeneralization errors if the rational set of non-examples is not provided, and misconception errors (both over- and undergeneralization errors) if neither the rational set of examples or non-examples is provided. Based on these predictions and the other relevant literature reviewed earlier, the following results were hypothesized for Experiment I;

(1) The ordering of the condition means on the dependent variables of correct classification of instances, recognition of definitions, and knowledge of relationships among concepts would be: Condition 1 > Condition 2 > Condition 3 > Condition 4.

(2) Providing only the rational set of examples would result in significantly more overgeneralization errors than providing the rational set of both examples and non-examples.

Subjects

Subjects were 96 sixth-grade students. They attended the same public elementary school in Menomonee Falls, Wisconsin as the subjects who participated in the pilot study. Subjects for Experiment I comprised the four sixth-grade classes in the school which had not taken part in the pilot. The subjects were stratified into three groups on the basis of reading achievement scores on the Iowa Tests of Basic Skills which was administered during the fall of the school year. Seventeen subjects were lost due to absence or because they had no reading achievement score. The final sample consisted of 79 subjects.

Materials

Lessons. Three versions of the basic series of lessons used in the pilot study and three placebo lessons were used in Experiment I. The three versions of the basic lessons varied according to the number of teaching examples and non-examples of the concepts which were presented. This variation in the lessons constituted the experimental treatments. In all cases the same instances which had been employed in the pilot study lessons were used. When fewer than the number of instances used in the pilot study were needed, that number was randomly selected from those used in the pilot. The placebo lessons dealt with the environmental concepts of population, habitat and community. The content of the lessons for each condition is presented in Table 1.

Tests. Test I-IV (with minor revisions based on the results of the pilot study) and Test V were used in Experiment I.

Table 1

Content of Lessons for Each Experimental Condition Experiment I

Treatment Lessons	Treatment Conditions			Control
	1	2	3	4
I Bilateral Symmetry	rational set of examples and non-examples	rational set of examples	Two examples	placebo lesson (population)
II Rotational Symmetry	rational set of examples and non-examples	rational set of examples	Two examples	placebo lesson (habitat)
III Translational Symmetry	rational set of examples and non-examples	rational set of examples	Two examples	placebo lesson (community)

Procedure

Prior to experimentation, subjects were randomly assigned within stratification levels to a treatment condition or the control group. Sets of the three concept lessons for each condition were then prepackaged and labeled with the subjects' names to ensure that each subject received the lessons appropriate to the experimental condition to which he had been assigned.

The materials were administered in two experimental sessions. At the beginning of the first session general instructions (Appendix C) concerning the purpose of the study and the procedures to be followed were read to the subjects, and the list of difficult words was reviewed in a manner similar to that used in the instance probability analysis. Subjects were then instructed to begin reading Lesson 1 and to raise their hands when they had finished. As each subject finished, an experimenter collected Lesson 1 and passed out Test 1. When subjects finished Test I, they handed it to an experimenter and began reading Lesson II. The entire procedure was then repeated for Lesson II and Test II, and then for Lesson III and Test III. When all of the subjects had finished Test III, Test IV was distributed.

Subjects were permitted to work on the lessons and tests for as long as they wished. The time needed ranged from 35 to 45 minutes. While the subjects were working the experimenter answered questions dealing with the pronunciation of words and also clarified directions.

The second experimental session took place exactly two weeks after the initial session. At this time Test V was administered. The experimenter first read a set of brief instructions (Appendix D) and then allowed the subjects to work through the test at their own rate. Time needed to complete Test V ranged from 10 to 20 minutes.

Experimental Design

The experiment employed a 3 x 4 factorial design with three levels of reading achievement and four types of lessons. The number of subjects in each cell of the design is shown in Table 2.

Experiment II

As discussed previously, studies by Anderson and Kulhavy (1972) and Merrill and Tennyson (1971) have shown that the presentation of a concept definition facilitates learning. The study by Frayer (1970) suggests further that any potential differences in concept attainment due to the number of instances provided are negated if a concept definition is also given. The purpose of Experiment II was to determine specifically whether the number of teaching instances presented has an effect on concept learning if a definition of the concept is also provided. It was essentially a replication of Experiment I with

Table 2

Number of Subjects in Each Cell

of Experimental Design

Experiment I

Stratification Levels	Conditions			
	1	2	3	4
1 (low scores)	7	6	7	7
2 (medium scores)	7	6	6	8
3 (high scores)	6	7	7	5
Totals	20	19	20	20

the addition of concept definitions in each condition. The dependent measures used to assess concept learning were identical to those used in Experiment I. The experimental conditions for Experiment II were:

> Condition 1: Concept lessons containing the rational set of both examples and non-examples with definitions of the concepts.
>
> Condition 2: Concept lessons containing the rational set of examples with definitions of the concepts.
>
> Condition 3: Concept lessons containing two examples with definitions of the concepts.
>
> Condition 4: Control condition: placebo lessons.

Based primarily on Frayer's findings, the following results were hypothesized for Experiment II:

(1) Condition 1, Condition 2, and Condition 3 would each result in significantly better performance than Condition 4 on the dependent variables of correct classification of instances, recognition of definitions, and knowledge of relationships among concepts.

(2) There would be no significant differences among Conditions 1, 2, and 3 on any of the five dependent variables.

Subjects

The initial sample consisted of 118 sixth-grade students from four public elementary schools in a rural Midwestern community. The subjects were divided into three groups on the basis of their reading achievement scores on the Iowa Tests of Basic Skills which was administered during the fall of the school year. Seven subjects were dropped from the analysis due to absence or because they lacked a reading achievement score. The final sample consisted of 111 subjects.

Materials

Lessons. The lessons consisted of three sets of treatment lessons and one set of three placebo lessons. The treatment lessons were identical to those used in Experiment I with the exception that they contained definitions of the concepts. Variations among the sets of lessons due to differences in the number of teaching instances constituted the experimental treatments. The three placebo lessons dealt with the concepts of number system, Roman numerals, and geometry. The content of the lessons for each condition is presented in Table 3.

Tests. Tests I-V were used in Experiment II. The content of the tests was identical to the content of the tests used in Experiment I with the exception of one item on Test III which was changed slightly to enhance clarity.

Table 3

Content of Lessons for Each
Experimental Condition
Experiment II

Treatment Lessons	Treatment Conditions			Control
	1	2	3	4
I Bilateral Symmetry	rational set of examples and non-examples with concept definition	rational set of examples with concept definition	two examples with concept definition	placebo lesson (number systems)
II Rotational Symmetry	rational set of examples and non-examples with concept definition	rational set of examples with concept definition	two examples with concept definition	placebo lesson (Roman Numerals)
III Translational Symmetry	rational set of examples and non-examples with concept definition	rational set of examples with concept definition	two examples with concept definition	placebo lesson (geometry)

Procedure

As in Experiment I the subjects were randomly assigned within stratification levels to a treatment condition or the control group. Prepackaged sets of the three lessons for each condition were then labeled with the subjects' names to insure that each subject received the lessons appropriate to the experimental condition to which he had been assigned.

The procedure followed in administering the materials was essentially identical to the procedure followed in Experiment I. The materials were again administered in two experimental sessions. However, to simplify the experimenters' task of simultaneously collecting, distributing, and organizing materials during the initial experimental session, all tests were collected simultaneously at the end of the session. Subjects were not permitted to look back at tests which they had completed. Time needed to complete the materials administered in the initial session ranged from 35 to 50 minutes. Time needed to complete the retention test (Test V) administered in the second session ranged from 10 to 20 minutes.

Experimental Design

The experiment employed a 3 x 4 factorial design with three levels of reading achievement and four types of lessons. The number of subjects in each cell of the design is shown in Table 4.

Table 4

Number of Subjects in Each Cell
of Experimental Design
Experiment II

Stratification Levels	Conditions			
	1	2	3	4
1 (low scores)	8	9	7	10
2 (medium scores)	10	9	10	9
3 (high scores)	10	8	9	10
Totals	28	26	28	29

Experiment III

The primary purpose of Experiment III was to determine the effect on concept learning of emphasizing the relevant attributes of the concept when both a definition of the concept and the rational set of teaching examples and non-examples are also presented. The dependent measures used to assess concept learning were identical to those used in Experiments I and II. The experimental conditions were:

Condition 1: Concept lessons containing the rational set of both examples and non-examples.

Condition 2: Concept lessons containing the rational set of both examples and non-examples with concept definitions.

Condition 3: Concept lessons containing the rational set of both examples and non-examples with concept definitions and emphasis of relevant attributes.

Condition 4: Control condition: placebo lessons.

The use of verbal cues to draw attention to relevant concept attributes has consistently been shown to facilitate concept learning (Frayer, 1970; Gelfand, 1958; Remstad, 1969; Wittrock, Keislar, & Stern, 1964). Additionally, studies by Anderson and Kulhavy (1972) and Merrill and Tennyson (1971) indicate that the presentation of a definition is also a powerful instructional variable. Based on the findings of these studies, the following results were hypothesized for Experiment III:

(1) The ordering of the condition means on the dependent variables of correct classification of instances, recognition of definitions, and knowledge of relationships among concepts would be: Condition 3 > Condition 2 > Condition 1 > Condition 4.

Subjects

The original sample consisted of 102 sixth-grade students. They attended four public elementary schools located in the same rural Midwestern community in which Experiment II was conducted. The subjects were initially divided into three stratification levels on the basis of their reading achievement scores on the Iowa Tests of Basis Skills which was administered during the beginnning of the school year. Two subjects were dropped from the analysis due to absence or because they lacked a reading achievement score. The final sample consisted of 100 subjects.

Materials

Lessons. The lessons consisted of three sets of treatment lessons and one set of three placebo lessons. The sets of treatment lessons were all versions of the basic lessons used in the pilot study. Variations in these lessons constituted the experimental treatments. The lessons for Condition 3 contained the rational set of positive and negative instances, concept definitions, and emphasis of relevant attribute values. These lessons were essentially identical to those used in the pilot. Lessons for Condition 2 contained the rational set of positive and negative instances and concept definitions. Lessons for Condition 1 contained only the rational set of positive and negative instances. The placebo lessons read by the control group were identical to those used in Experiment II. The content of the lessons for each condition is outlined in Table 5.

Tests. Tests I-V were used in Experiment III. The content of the tests was identical to the content of the tests used in Experiment II.

Procedure

As in Experiments II and III, the subjects were randomly assigned within stratification levels to a treatment condition or the control group prior to experimentation. Prepackaged sets of the three lessons for each of the four conditions were then labeled with the subjects' names to insure that each subject received the lessons appropriate to the experimental condition to which he had been assigned.

The procedure followed in administering the experimental materials was identical to that followed in Experiment II. Time needed to complete the materials administered in the first experimental session ranged from 35 to 50 minutes. Time needed to complete the retention test (Test V) administered in the second experimental session ranged from 10 to 20 minutes.

Table 5

Content of Lessons for Each
Experimental Condition
Experiment III

Treatment Lessons	Treatment Conditions			Control
	1	2	3	4
I Bilateral Symmetry	rational set of examples and non-examples	rational set of examples and non-examples with concept definition	rational set of examples and non-examples with concept definition and emphasis	placebo lesson (number systems)
II Rotational Symmetry	rational set of examples and non-examples	rational set of examples and non-examples with concept definition	rational set of examples and non-examples with concept definition and emphasis	placebo lesson (Roman numerals)
IIII Translational Symmetry	rational set of examples and non-examples	rational set of examples and non-examples with concept definition	rational set of examples and non-examples with concept definition and emphasis	placebo lesson (geometry)

Experimental Design

The experiment employed a 3 x 4 factorial design with three levels of reading achievement and four types of lessons. The number of subjects in each cell of the design is shown in Table 6.

Table 6

Number of Subjects in Each Cell
of Experimental Design
Experiment III

Stratification Levels	Conditions			
	1	2	3	4
1 (low scores)	8	12	8	7
2 (medium scores)	8	8	6	8
3 (high scores)	8	8	10	9
Totals	24	28	24	24

Chapter V

RESULTS

Standardization Procedure for Experiments I, II, and III

An overall score on each of the dependent variables (correct classification of instances, recognition of definitions, knowledge of relationships among concepts, overgeneralization errors, and undergeneralization errors) was calculated for every subject from Tests I-IV. These scores constituted measures of immediate concept acquisition. For the variables of correct classification of instances, recognition of definitions, and over- and undergeneralization errors, calculation of an overall score necessitated combining scores across tests. To eliminate the possibility of differences due to individual test difficulty or variability, scores for these variables were first converted to standard scores within each test and then summed across tests to yield one score for each variable.

The results of the retention test (Test V) were analyzed separately from the results of Tests I-IV. The first three sections of Test V, each of which dealt with the identification of instances of one of the three geometric concepts (bilateral symmetry -- Section A; rotational symmetry -- Section B; and translational symmetry -- Section C), were first converted to standard scores within concept and then combined to form one retention measure for the variables of correct classification of instances, over- and undergeneralization errors.

The means (raw scores) and standard deviations for each part of Tests I-V are included (for each experiment) as Appendix E. They are presented by condition and stratification level.

Experiment I

The two specific hypotheses tested in Experiment I were: (1) that the ordering of the condition means on the dependent variables of correct classification of instances, recognition of definitions, and knowledge of relationships among concepts would be: Condition 1 (lessons containing the rational set of both examples and non-examples) > Condition 2 (lessons containing the rational set of examples only) > Condition 3 (lessons containing just two examples) > Condition 4 (control); and (2) that providing the rational set of examples would result in significantly more overgeneralization errors than providing the rational set of both examples and non-examples.

A 3 X 4 analysis of variance (three levels of stratification and four levels of type of lesson) was performed on the data for each of the five dependent variables. When the main effect for condition was found to be significant, all pairwise comparisons among the means were tested using the method developed by Tukey (Meyers, 1966). As the cell sizes were unequal, an approximate critical value based on the harmonic mean was used in the comparisons. All comparisons were tested at the .05 significance level. Although pairwise Tukey comparisons were not carried out in cases where the analysis of variance showed no significant condition effects, the ordering of the condition means will be reported for the reader's information. The means for each condition by stratification level and the results of the analyses of variance are presented in Tables 7 and 8 for immediate acquisition and in Tables 9 and 10 for retention.

Correct Classification of Instances

Immediate Acquisition. The results of the analysis of variance showed a signficant main effect for both conditions ($F=3.2099$, $p<.0285$) and stratification levels ($F=13.6525$, $p<.0001$).* The condition x stratification interaction was not signficant.

The ordering of the condition means was: Condition 1 > Condition 3 > Condition 2 > Condition 4. This was not in the predicted direction. However, using Tukey's procedure, only the difference between the mean for subjects who read lessons containing the rational set of both examples and non-examples (Condition 1) and the mean for the control subjects (Condition 4) was found to be signficant ($p<.05$). Thus, providing the rational set of both examples and non-examples was necessary for concept learning to occur as measured by the subjects' ability to identify new instances.

Retention. The results of the analysis of variance showed that the main effect for stratification was significant ($F=15.1087$, $p<.0001$). However, neither the condition effect nor the interaction was significant. An examination of the condition means shows the ordering was not in the predicted direction: Condition 2 > Condition 1 > Condition 3 > Condition 4.

*The means for stratification level showed that good readers performed better than poorer readers. This finding was consistent across all experiments whenever a significant stratification effect was found and will not be mentioned again.

Table 7

Experiment I
Condition Means by Stratification Level
for all Dependent Variables
Immediate Acquisition

Dependent Variables		1	2	3	4
		Condition Means			
1. Correct Class.	Strat.Levels 1	0.6696	-0.7234	-1.4107	-1.7504
of Instances	2	-0.6381	-0.9709	-0.5156	-0.4763
(z scores)	3	2.9194	0.5447	2.2961	0.3167
MSE = 3.1199	Means over strat.levels	0.8869	-0.3344	0.1552	-0.7240
2. Recog. of	Strat.Levels 1	0.4008	-1.6575	-0.0453	-1.1521
Defin.	2	-1.0408	0.4857	-1.5988	- .1136
(z scores)	3	2.2262	0.9527	0.7034	1.1172
MSE = 4.0347	Means over strat.levels	0.4439	-0.0190	-0.2493	-0.1694
3. Knowl. of	Strat.levels 1	2.4286	2.1667	1.4286	1.2857
Relat.	2	2.0000	2.0000	1.6667	1.7500
	3	3.0000	2.2857	3.1429	2.2000
MSE = 1.3374	Means over strat.levels	2.4500	2.1579	2.1000	1.7000
4. Overgen.	Strat.Levels 1	0.6039	0.8126	1.0047	1.0217
Errors	2	-0.1076	1.7898	-0.2689	-0.1292
(z scores)	3	-1.8804	-0.2452	-2.5064	-0.0152
MSE = 3.7198	Means over strat.levels	-0.3904	0.7315	-0.6063	0.3021
5. Undergen.	Strat.Levels 1	-1.3925	-0.0646	1.2918	1.6185
Errors	2	1.1812	-0.6322	1.0933	0.9717
(z scores)	3	-2.5231	-0.4227	-1.2103	-0.4954
MSE = 3.8378	Means over strat.levels	-0.8309	-0.3758	0.3566	0.8313

Table 8

Experiment I. Analyses of Variance for

Immediate Acquisition Data

Variable	Source	df	MS	F	p<
Correct Class. of Instances	Condition	3	10.0144	3.2099	.0285
	Strat.	2	42.5944	13.6525	.0001
	C x S	6	4.5151	1.4472	.2100
	Error	67	3.1199		
Recog. of Defin.	Condition	3	2.5392	.6293	.5987
	Strat.	2	28.4872	7.0605	.0017
	C x S	6	6.0082	1.4891	.1954
	Error	67	4.0347		
Knowl. of Relat.	Condition	3	1.7583	1.3147	.2770
	Strat.	2	5.6093	4.1940	.0193
	C x S	6	1.0434	.7801	.5885
	Error	67	1.3374		
Overgen. Errors	Condition	3	7.4639	2.0065	.1214
	Strat.	2	30.3651	8.1632	.0007
	C x S	6	3.8747	1.0417	.4066
	Error	67	3.7198		
Undergen. Errors	Condition	3	10.9516	2.8536	.0437
	Strat.	2	23.9371	6.2372	.0033
	C x S	6	6.7843	1.7678	.1192
	Error	67	3.8378		

Note.--Analyses are based on unequal n's.

Table 9

Experiment I. Condition Means by Stratification Level for all Dependent Variables Retention

Dependent Variables		Condition Means 1	2	3	4
1. Correct Class. of Instances (z scores)	Strat.Levels 1	0.4432	-0.1813	-1.0201	-1.4638
	2	-1.0830	-0.4057	-0.7744	-1.5256
	3	1.3829	2.2446	1.9182	0.9622
MSE = 3.2130	Means over strat.levels	0.1910	0.6416	0.0820	-0.8820
2. Recog. of Defin.	Strat.Levels 1	0.4286	0.5000	0.1429	0.4286
	2	0.5714	1.1667	0.5000	1.000
	3	0.6667	1.5714	0.8571	0.8000
MSE = 0.6647	Means over strat.levels	0.5500	1.1053	0.5000	0.7500
4. Overgen. Errors (z scores)	Strat.Levels 1	-0.1826	0.9102	-0.1066	1.6540
	2	0.2710	0.7870	0.3962	0.5286
	3	-0.5687	-1.6595	-1.0590	-1.1592
MSE = 4.6025	Means over strat.levels	-0.1396	-0.0754	-0.2891	0.5005
5. Undergen. Errors (z scores)	Strat.Levels 1	-0.7740	-0.7895	0.6045	0.9272
	2	1.3382	0.1240	1.0337	1.8843
	3	-1.2873	-1.5565	-1.7510	-0.2159
MSE = 3.5291	Means over strat.levels	-0.1887	-0.7836	-0.0912	1.0243

Table 10

Experiment I. Analyses of Variance for Retention Data

Variable	Source	df	MS	F	p<
Correct Class. of Instances	Condition	3	5.2261	1.6265	.1915
	Strat.	2	48.5447	15.1087	.0001
	C x S	6	1.5534	.4835	.8185
	Error	67	3.2130		
Recog. of Defin.	Condition	3	1.3165	1.9805	.1253
	Strat.	2	2.4935	3.7510	.0286
	C x S	6	.2906	.4372	.8515
	Error	67	.6667		
Overgen. Errors	Condition	3	2.3936	.5201	.6700
	Strat.	2	22.1589	4.8145	.0112
	C x S	6	2.6885	.5841	.7418
	Error	67	4.6025		
Undergen. Errors	Condition	3	11.1760	3.1668	.0300
	Strat.	2	34.5616	9.7934	.0002
	C x S	6	1.6498	.4675	.8301
	Error	67	3.5291		

Note.--Analyses are based on unequal *n*'s.

Recognition of Definitions

Immediate Acquisition and Retention. The results of the analysis of variance for both immediate acquisition and retention showed that the only significant effect was due to stratification (immediate acquisition: $F=7.0605$, $p<.0017$; retention: $F=3.7510$, $p<.0286$). In neither case was the ordering of the condition means in the predicted direction (immediate acquisition: Condition 1 > Condition 2 > Condition 4 > Condition 3; retention: Condition 2 > Condition 4 > Condition 1 > Condition 3).

Knowledge of Relationships among Concepts

Again, the only significant effect was due to stratification ($F=4.1940$, $p<.0193$). However, the ordering of the condition means was in the predicted direction.

Overgeneralization Classification Errors

Immediate Acquisition. The results of the analysis of variance showed that there were no signficant condition or interaction effects, but that the main effect for stratification was signficant ($F=8.1632$, $p<.0007$). The ordering of the condition means was: Condition 2 > Condition 4 > Condition 1 > Condition 3.

Retention. Again, neither condition nor interaction effects were found to be significant, but the main effect for stratification was significant ($F=4.8145$, $p<.0112$). The ordering of the condition means was: Condition 4 > Condition 2 > Condition 1 > Condition 3.

Undergeneralization Classification Errors

Immediate Acquisition. The results of the analysis of varianceiance showed a significant main effect for both conditions (F= 2.8536, p<.0437) and stratification levels (F=6.2372, p<.0033). The condition x stratification effect was not significant.

Pairwise comparisons of the condition means showed that the only significant difference was between the control group and subjects who read lessons containing the rational set of examples and non-examples (Condition 1; p<.05), with subjects in the control group making significantly more undergeneralization errors. The ordering of the means was: Condition 4 > Condition 3 > Condition 2 > Condition 1.

Retention. The main effects for both condition and stratification level were found to be significant (F=3.1668, p<.03, and F=9.7934, p<.0002, respectively). The interaction effect was not found to be significant.

The ordering of the condition means was: Condition 4 > Condition 3 > Condition 1 > Condition 2. The only significant (p<.05) difference was found to be between the control group (Condition 4) and subjects who read lessons containing the rational set of examples (Condition 2).

Experiment II

The two specific hypotheses tested in Experiment II were: (1) that Condition 1 (lessons containing the rational set of both examples and non-examples with definitions), Condition 2 (lessons containing the rational set of examples with definitions), and Condition 3 (lessons containing two examples with definitions), would each result in significantly better performance on the dependent variables of correct classification of instances, recognition of definitions, and knowledge of relationships among concepts than Condition 4 (control); and (2) that due to the addition of concept definitions to the treatment conditions, there would be no significant differences among Conditions 1, 2, and 3 on any of the five dependent variables.

A 3 X 4 analysis of variance (with three levels of stratification and four levels of type of lesson) was performed on the data for each of the dependent variables for immediate acquisition and retention. When the main effect for condition was found to be signficant, all pairwise comparisons among the means were tested using Tukey's method. An approximate critical value based on the harmonic mean was used in the comparisons as the cell sizes were unequal. Each comparison was tested at the .05 significance level. The ordering of the condition means will again be reported in all cases for the reader's information. The means for each condition by stratification level and the results of the analyses of variance are reported in Tables 11 and 12 for immediate acquisition and Tables 13 and 14 for retention.

Correct Classification of Instances

Immediate Acquisition. The results of the analysis of variance showed that the main effect for both conditions and stratification levels was significant ($F=15.0053$, $p<.0001$, and $F=18.3073$, $p<.0001$, respectively). The interaction was also

Table 11

Experiment II. Condition Means by Stratification Level for all Dependent Variables Immediate Acquisition

Dependent Variables		Condition Means 1	2	3	4
1. Correct Class.	Strat.Levels 1	-1.5359	-1.0110	-1.1709	-1.6984
of Instances	2	0.7877	1.6394	0.8544	2.0396
(z scores)	3	1.3021	2.0170	2.7601	1.7822
MSE = 3.0517	Means over strat.levels	0.3075	0.8350	0.8160	-1.8331
2. Recog. of	Strat.Levels 1	-1.7716	-0.9923	-0.3034	-2.7100
Defin.	2	0.5974	0.3460	1.1549	-2.0595
(z scores)	3	2.1918	1.6760	3.3397	-1.3737
MSE = 4.9669	Means over strat.levels	0.4900	0.2920	1.3884	-2.0783
3. Knowl. of	Strat.Levels 1	1.1250	1.6667	1.7778	1.0000
Relat.	2	2.8000	2.3333	2.5000	2.0000
	3	3.9000	4.0000	4.7778	2.2000
MSE = 2.1069	Means over strat.levels	2.7143	2.6154	3.0000	1.7241
4. Overgen.	Strat.Levels 1	1.0000	2.0225	0.8279	1.4018
Errors	2	-0.7623	-0.6191	-0.8772	1.4338
(z scores)	3	-1.6436	-1.2810	-2.0517	0.6538
MSE = 3.4756	Means over strat.levels	-0.5735	0.0916	-0.7067	1.1538
5. Undergen.	Strat.Levels 1	1.5295	-0.2773	1.0065	1.2870
Errors	2	-0.5669	-1.8374	-0.4598	1.8408
(z scores)	3	-0.8248	-1.8392	-2.1675	2.1038
MSE = 3.5439	Means over strat.levels	-0.0600	-1.2979	-0.5374	1.7405

Table 12

Experiment II. Analyses of Variance for

Immediate Acquisition Data

Variable	Source	df	MS	F	p<
Correct Class. of Instances	Condition	3	45.7919	15.0053	.0001
	Strat.	2	55.8686	18.3073	.0001
	C x S	6	8.7555	2.8691	.0128
	Error	99	3.0517		
Recog. of Defin.	Condition	3	61.8541	12.4532	.0001
	Strat.	2	77.4973	15.6027	.0001
	C x S	6	3.3561	.6757	.6696
	Error	99	4.9669		
Knowl. of Relat.	Condition	3	8.5190	4.0433	.0094
	Strat.	2	49.3717	23.4330	.0001
	C x S	6	2.5786	1.2239	.3006
	Error	99	2.1069		
Overgen. Errors	Condition	3	20.6719	5.9477	.0010
	Strat.	2	50.9104	14.6478	.0001
	C x S	6	4.0921	1.1774	.3246
	Error	99	3.4756		
Undergen. Errors	Condition	3	46.6131	13.1532	.0001
	Strat.	2	21.6840	6.1187	.0032
	C x S	6	8.0826	2.2807	.0420
	Error	99	3.5439		

Note.--Analyses are based on unequal *n*'s.

Table 13

Experiment II. Condition Means by Stratification Level for all Dependent Variables Retention

Dependent Variables		Condition Means 1	2	3	4
1. Correct Class.	Strat.Levels 1	-0.0913	-0.5708	-1.2288	-2.4307
of Instances	2	0.6491	0.5951	0.9133	-2.2597
(z scores)	3	1.9030	2.2800	2.2800	-1.0124
MSE = 3.8219	Means over strat.levels	0.8854	0.4381	0.6641	-1.8885
2. Recog. of	Strat.Levels 1	0.7500	1.2222	0.7778	0.6000
Defin.	2	1.9000	1.6667	1.7000	0.5556
	3	2.1000	1.7500	2.7778	1.3000
MSE = 0.8918	Means over strat.levels	1.6429	1.5385	1.7500	0.8276
4. Overgen.	Strat.Levels 1	0.2559	0.4626	1.3024	1.9203
Errors	2	-0.5182	0.1049	-1.0715	2.0494
(z scores)	3	-1.6280	-1.1691	-1.9426	0.2492
MSE = 5.0159	Means over strat.levels	-0.6934	-0.1633	-0.5885	1.3841
5. Undergen.	Strat.Levels 1	-0.0591	0.2422	0.6512	1.8190
Errors	2	-0.4467	-0.9763	-0.3419	1.3815
(z scores)	3	-1.2952	-0.9903	-1.5936	1.3697
MSE = 3.6687	Means over strat.levels	-0.6390	-0.5588	-0.4250	1.5283

Table 14

Experiment II. Analyses of Variance for

Retention Data

Variable	Source	df	MS	F	p<
Correct Class. of Instances	Condition	3	46.9763	12.2913	.0001
	Strat.	2	44.7842	11.7177	.0001
	C x S	6	2.4055	.6294	.7064
	Error	99	3.8219		
Recog. of Defin.	Condition	3	4.8976	5.4918	.0016
	Strat.	2	11.8805	13.3218	.0001
	C x S	6	1.4300	1.6035	.1541
	Error	99	.8918		
Overgen. Errors	Condition	3	26.4693	5.2770	.0021
	Strat.	2	40.7095	8.1160	.0006
	C x S	6	2.9752	.5932	.7352
	Error	99	5.0195		
Undergen. Errors	Condition	3	30.7818	8.3903	.0001
	Strat.	2	15.0148	4.0926	.0196
	C x S	6	1.7071	.4653	.8325
	Error	99	3.6687		

Note.--Analyses are based on unequal _n_'s.

significant (F=2.8691, p<.0128). An examination of the means in Table 11 shows that the interaction effect was significant because the ordering of the means by stratification level was exactly the same in each treatment condition (level 3 > level 2 > level 1) but differed in the control condition (level 2 > level 3 > level 1).

The ordering of the condition means was: Condition 2 > Condition 3 > Condition 1 > Condition 4. Tukey comparisons showed that each of Conditions 1, 2 and 3 were significantly different from the control group (p<.05). No significant differences were found between the treatment means. Thus, both Hypotheses 1 and 2 were confirmed, indicating that providing definitions with teaching instances promotes concept attainment, and that the presence of definitions negates any effects due to the number of teaching instances provided.

Retention. The condition effect and the stratification effect were significant (F=12.2913, p<.0001, and F=11.7177, p<.0001, respectively). The interaction effect was not significant.

The ordering of the condition means was: Condition 1 > Condition 3 > Condition 2 > Condition 4. As on immediate acquisition, each of the treatment conditions was found to be significantly different from the control condition (p<.05 in all cases). Furthermore, no significant differences were found among the treatment conditions. Thus these results also confirmed Hypotheses 1 and 2.

Recognition of Definitions

Immediate Acquisition. Main effects for condition and stratification level were significant (F=12.4532, p<.0001, and

_F_=15.6027, _p_<.0001, respectively). The condition x stratification interaction was not significant.

The ordering of the condition means was: Condition 3 > Condition 1 > Condition 2 > Condition 4. Pairwise comparisons among the means showed that both of the hypotheses were again confirmed. No significant differences were found among the treatment conditions, but each treatment condition was found to be significantly different from the control (_p_<.05 in each case).

Retention. The analysis of variance showed that the condition effect and the stratification effect were both significant (_F_=5.4918, _p_<.0016, and _F_=13.3218, _p_<.0001, respectively). The interaction effect was not significant.

The ordering of the condition means was: Condition 3 > Condition 1 > Condition 2 > Condition 4. Again, each treatment condition was found to be significantly different from the control condition (_p_<.05 in all cases) but no significant differences were found among the treatment conditions.

Knowledge of Relationships among Concepts

The main effects for condition and stratification level were signficant (_F_=4.0433, _p_<.0094, and _F_=23.4330, _p_<.0001, respectively). The condition x stratification interaction was not significant.

Pairwise comparisons among the means showed that the only significant difference was between Condition 3 (two examples with definitions) and the control group. The ordering of the means was: Condition 3 > Condition 1 > Condition 2 > Condition 4.

Overgeneralization Classification Errors

Immediate Acquisition. The results of the analysis of variance showed that condition and stratification effects were signficant (*F*=5.9477, *p*<.0010, and *F*=14.6478, *p*<.0001, respectively). The interaction effect was not significant.

The ordering of the condition means was: Condition 4 > Condition 2 > Condition 1 > Condition 3. The differences between Condition 4 (control) and Condition 1 (rational set of examples and non-examples with definitions), and between Condition 4 and Condition 3 (two examples with definitions), were both found to be significant (in both cases *p*<.05).

Retention. The results of the analysis of variance showed significant main effects for both conditions (*F*=5.2770, *p*<.0021) and stratification levels (*F*=8.1160, *p*<.0006). The interaction effect was not significant.

The ordering of the condition means was: Condition 4 > Condition 2 > Condition 3 > Condition 1. Again, the differences between Conditions 4 and 1 and between Conditions 4 and 3 were significant (*p*<.05 in both cases).

Undergeneralization Classification Errors

Immediate Acquisition. The main effects for condition and stratification level were found to be signficant (*F*=13.1532, *p*<.0001, and *F*=6.1187, *p*<.0032, respectively). The interaction effect was also significant (*F*=2.2807, *p*<.0420). An examination of the means in Table 11 shows that this was caused by the fact that while in each treatment condition subjects in the highest stratification level (3) made fewer errors than subjects in the middle level (2) who in turn made fewer errors than subjects in the lowest level (1), this order was exactly reversed in the control condition.

The ordering of the condition means was: Condition 4 > Condition 1 > Condition 3 > Condition 2. Significant differences were found between the control and each of the treatment conditions (*p*<.05 in all cases).

Retention. Both the condition and the stratification effect were signficant (*F*=8.3909, *p*<.0001, and *F*=4.0926, *p*<.0196, respectively). The interaction was not significant.

The ordering of the condition means was: Condition 4 > Condition 3 > Condition 2 > Condition 1. Again, significant differences were found between the control and each of the treatment conditions (*p*<.05 in all cases).

Experiment III

The predicted ordering of the condition means in Experiment III on the dependent variables of correct classification of instances, recognition of definitions, and knowledge of relationships among concepts was: Condition 3 (lessons containing the rational set of examples and non-examples with both concept definitions and emphasis) > Condition 2 (lessons containing the rational set of examples and non-examples with just definitions) > Condition 1 (lessons containing the rational set of examples and non-examples only) > Condition 4 (control). The direction of the means on the variables of over- and undergeneralization of classification errors was not hypothesized.

A 3 X 4 analysis of variance (with three levels of stratification and four levels of type of lesson) was performed on the data for each dependent variable. Tukey tests were used to test all pairwise comparisons among the condition means when the main effect for condition was found to be significant. An approximate critical value based on the harmonic mean was used in the comparisons due to the unequal cell sizes. All comparisons were tested at the .05 level of significance. The ordering of the condition means will be reported in all cases for the reader's information even when there was no significant condition effect. The means for each condition by stratification level and the results of the analyses of variance are presented in Tables 15 and 16 for immediate acquisition and in Tables 17 and 18 for retention.

Correct Classification of Instances

Immediate Acquisition. The results of the analysis of variance showed significant main effects for both conditions (F=17.6969, p<.0001) and stratification level (F=17.0721, p<.0001). The interaction was not significant.

The ordering of the condition means was in the predicted direction. Tukey comparisons showed that both Condition 3 (rational set of examples and non-examples with definitions and emphasis) and Condition 2 (rational set of examples and non-examples with definitions) differed from the control condition (p<.05). Additionally, each of Conditions 3 and 2 were found to be signficantly different from Condition 1 (rational set of examples and non-examples), but not from each other. Thus, providing the rational set of examples and non-examples

Table 15

Experiment III
Condition Means by Stratification Level
for all Dependent Variables
Immediate Acquisition

Dependent Variables		Condition Means 1	2	3	4
1. Correct Class.	Strat.Levels 1	-1.5492	0.1176	-0.6586	-3.3319
of Instances	2	-5.4830	0.8020	1.8687	-1.5375
(z scores)	3	0.3962	2.2454	2.2454	-0.8503
MSE = 3.0386	Means over strat.levels	-0.5671	1.0175	1.1832	-1.8032
2. Recog. of	Strat.Levels 1	-2.0752	-0.5548	1.0784	-2.9326
Defin.	2	-2.0832	1.5454	0.6851	-1.1097
(z scores)	3	-0.8742	3.0165	3.1370	-0.4756
MSE = 6.0355	Means over strat.levels	-1.6775	1.0657	1.8378	-1.4036
3. Knowl. of	Strat.Levels 1	2.1250	2.2500	2.8750	1.4286
Relat.	2	2.2500	3.5000	3.0000	1.7500
	3	3.0000	4.3750	4.5000	3.4444
MSE = 3.1740	Means over strat.levels	2.4583	3.2143	3.5833	2.2917
4. Overgen.	Strat.Levels 1	1.6563	-0.0580	0.7282	2.4552
Errors	2	0.6883	-0.3707	-1.9164	0.7123
(z scores)	3	-0.2709	-1.8331	-1.9302	0.4252
MSE = 3.2177	Means over strat.levels	0.6912	-0.6545	-1.0406	1.1130
5. Undergen.	Strat.Levels 1	0.9250	-0.1105	0.3908	3.0324
Errors	2	0.2878	-0.9225	-1.1870	1.8555
(z scores)	3	-0.3966	-2.4443	-1.8830	0.9428
MSE = 2.9147	Means over strat.levels	0.2721	-1.0093	-0.9510	1.8565

Table 16

Experiment III. Analyses of Variance for

Immediate Acquisition Data

Variable	Source	df	MS	F	p<
Correct Class. of Instances	Condition	3	53.7745	17.6969	.0001
	Strat.	2	51.8760	17.0721	.0001
	C x S	6	1.5273	.5026	.8049
	Error	88	3.0386		
Recog. of Defin.	Condition	3	76.9422	12.7483	.0001
	Strat.	2	46.9405	7.7774	.0008
	C x S	6	4.3147	.7149	.6386
	Error	88	6.0355		
Knowl. of Relat.	Condition	3	9.7668	3.0771	.0317
	Strat.	2	25.0786	7.9013	.0007
	C x S	6	1.1852	.3734	.8941
	Error	88	3.1740		
Overgen. Errors	Condition	3	26.3936	8.2025	.0001
	Strat.	2	37.0375	11.5104	.0001
	C x S	6	2.0842	.6477	.6918
	Error	88	3.2177		
Undergen. Errors	Condition	3	44.9092	15.4079	.0001
	Strat.	2	34.8091	11.9427	.0001
	C x S	6	.7025	.2410	.9617
	Error	88	2.9147		

Note.--Analyses are based on unequal *n*'s.

Table 17

Experiment III. Condition Means by Stratification Level for all Dependent Variables Retention

Dependent Variables		Condition Means 1	2	3	4
1. Correct class.	Strat.Levels 1	-2.0708	-0.3240	-1.0376	-2.2570
of Instances	2	-4.3362	1.0801	0.7162	-1.0309
(z scores)	3	0.7363	2.2169	2.2990	-0.3652
MSE = 4.5626	Means over strat.levels	-0.5893	0.8032	0.7911	-1.1389
2. Recog. of	Strat.Levels 1	1.1250	1.4167	2.1250	1.0000
Defin.	2	0.8750	2.5000	1.5000	1.2500
	3	1.8750	2.3750	2.7000	1.4444
MSE = 1.2456	Means over strat.levels	1.2917	2.0000	2.2083	1.2500
4. Overgen.	Strat.Levels 1	1.4756	0.6648	1.0463	1.7349
Errors	2	0.6033	-0.4397	-1.2114	0.2155
(z scores)	3	-0.0863	-1.9468	-2.2130	0.2593
MSE = 4.7365	Means over strat.levels	0.6642	-0.3969	-0.8762	0.6751
5. Undergen.	Strat.Levels 1	1.9493	-0.0414	0.4440	2.0572
Errors	2	0.2064	-1.2746	0.0093	1.1824
(z scores)	3	-0.9578	-1.7736	-1.6076	0.4342
MSE = 4.4334	Means over strat.levels	0.3993	-0.8887	-0.5195	1.1570

Table 18

Experiment III. Analyses of Variance for

Retention Data

Variable	Source	df	MS	F	p<
Correct Class. of Instances	Condition	3	26.6783	5.8471	.0011
	Strat.	2	59.9450	13.1382	.0001
	C x S	6	.7903	.1732	.9834
	Error	88	4.5626		
Recog. of Defin.	Condition	3	5.7665	4.6294	.0048
	Strat.	2	4.5236	3.6316	.0306
	C x S	6	1.3473	1.0818	.3796
	Error	88	1.2456		
Overgen. Errors	Condition	3	14.7874	3.1220	.0300
	Strat.	2	44.5383	9.4032	.0002
	C x S	6	2.1175	.4471	.8453
	Error	88	4.7365		
Undergen. Errors	Condition	3	21.5149	4.8530	.0036
	Strat.	2	37.4479	8.4468	.0005
	C x S	6	1.1252	.2538	.9566
	Error	88	4.4334		

Note.--Analyses are based on unequal _n_'s.

alone did not result in concept attainment as measured by the subjects' ability to identify new instances, but providing the rational set of examples and non-examples with definitions, or with definitions plus emphasis, did result in concept learning. Moreover, the results also indicate that emphasis of relevant attributes does not significantly increase subjects' performance if a definition and the rational set of examples and non-examples are also provided.

Retention. Main effects for condition (F=5.8471, p<.0011) and stratification (F=13.1382, p<.0001) were again found to be significant. The interaction effect was not significant.

The ordering of the condition means was not in the hypothesized direction: Condition 2 > Condition 3 > Condition 1 > Condition 4. Significant differences were found only between Condition 2 and the control and Condition 3 and the control (p<.05 in both cases).

Recognition of Definitions

Immediate Acquisition. The results of the analysis of variance showed a significant main effect for both condition (F=12.7483, p<.0001) and stratification level (F=7.7774, p<.0008). The condition x stratification interaction was not significant.

The ordering of the condition means was: Condition 3 > Condition 2 > Condition 4 > Condition 1. Means for Condition 3 (rational set of examples and non-examples with definitions and emphasis) and Condition 2 (rational set of examples and non-examples with definitions) were each significantly different from the mean of the control group ($p<.05$ in both cases), and signficantly different from the mean for Condition 1 (rational set of examples and non-examples alone). Apparently, providing the rational set of examples and non-examples alone did not provide subjects with enough information to identify the concept definitions. However, subjects who had been given the definitions with the rational set of examples and non-examples could identify definitions more successfully than the control group. The addition of emphasis of relevant attributes to the rational set of examples and non-examples with definitions did not significantly improve performance.

Retention. Again, condition and stratification effects were signficant ($F=4.6294$, $p<.0048$, and $F=3.6316$, $p<.0306$, respectively). The interaction was not significant.

The ordering of the means was in the predicted direction. Significant differences were found between Conditions 3 and 4 ($p<.05$) and Conditions 3 and 1 ($p<.05$). It thus appears that only subjects who read lessons containing the rational set of examples and non-examples with definitions and emphasis could remember significantly more definitions than the control group on retention.

Knowledge of Relationships among Concepts

The main effects for condition ($F=3.0771$, $p<.0317$) and stratification ($F=7.9013$, $p<.0007$) were significant. However, none of the Tukey comparisons among the means was found to be significant, although the ordering of the condition means was in the hypothesized direction. The condition x stratification interaction effect was not significant.

Overgeneralization Classification Errors

Immediate Acquisition. Both condition and stratification effects were found to be significant (F=8.2025, p<.0001, and F=11.5104, p<.0001, respectively). The interaction effect was not found to be significant.

The ordering of the condition means was: Condition 4 > Condition 1 > Condition 2 > Condition 3. Significant differences were found between the means for Condition 2 (rational set plus definitions) and the control (p<.05), and Condition 3 (rational set plus definitions and emphasis) and the control (p<.05). In both cases the subjects in the control condition overgeneralized more than subjects in the two treatment conditions. A significant difference between Conditions 1 and 2, and 1 and 3, was also found (p<.05). Subjects in Condition 1 made significantly more overgeneralization errors than subjects in Conditions 2 or 3.

Retention. Significant condition (F=3.1220, p<.0300) and stratification (F=9.4032, p<.0002) effects were again found. The interaction effect was not significant.

None of the pairwise comparisons among the condition means was found to be significant. The ordering of the means was the same as that for immediate acquisition of overgeneralization classification errors.

Undergeneralization Classification Errors

Immediate Acquisition. The results of the analysis of variance showed that both the main effects for condition and stratification were significant (F=15.4079, p<.0001, and F=11.9427, p<.0001, respectively). Again, the interaction effect was not significant.

The direction of the means was: Condition 4 > Condition 1 > Condition 3 > Condition 2. Testing all pairwise comparisons between the condition means showed that the mean for each of the treatment conditions was significantly different from the mean for the control group (in all cases p<.05), and the mean for Condition 2 (rational set with definitions) was significantly different from the mean for Condition 1 (rational set only).

Retention. Significant differences were again found for the main effects of condition (F=4.8530, p<.0036) and stratification (F=8.4468, p<.0005). The interaction effect was not significant.

The direction of the means was: Condition 4 > Condition 1 > Condition 3 > Condition 2. Significant differences were found between Condition 2 and the control and Condition 3 and the control (p<.05 in both cases).

Chapter VI

DISCUSSION

The purpose of the present series of experiments was to determine the effect of various instructional variables on concept learning at the formal level. Three instructional variables were studied: number of teaching examples and non-examples, concept definition, and emphasis of relevant attributes.

Experiment I focused on the effects of the number of teaching examples and non-examples presented when the concepts used were not defined. Results showed that subjects who read lessons containing the rational set of both examples and non-examples were able to identify significantly more new instances on tests of immediate acquisition than subjects in the control group. However, neither subjects who read lessons containing the rational set of examples only nor subjects who read lessons containing two examples differed significantly in performance from the control group. Although there were no significant differences among the experimental conditions on the retention measure, the fact that there was a significant difference between subjects reading lessons with the rational set of both examples and non-examples and the control on immediate acquisition supports the views of both Markle and Tiemann (1969) and Merrill (1971) concerning the importance of non-examples in concept learning. Apparently, as Tennyson (1971) also found, non-examples function as a facilitative factor in concept learning when they are chosen according to rational criteria and used in a way which focuses the subject's attention on the relevant attributes of the concept.

Results of Experiment I also indicate that the number of examples provided when no non-examples were given was not a critical factor. Indeed, providing examples alone, even the rational set of examples, appears to have been singularly ineffective in promoting concept learning. Very likely this was due in part to the difficult nature of the concepts themselves, and the fact that the properties of the instances which were relevant concept attributes were not immediately obvious.

An analysis of the pattern of errors made in Experiment I lent support to Markle and Tiemann's theoretical position regarding classification errors. Although no differences between the treatment conditions on over- or undergeneralization were found to be statistically significant, subjects who read lessons containing the rational set of examples only did make more overgeneralization errors on immediate acquisition and retention than subjects who read lessons containing the rational set of examples and non-examples. This is precisely the result Markle and Tiemann have predicted. They theorize that without non-examples the student does not learn to discriminate the concept from other concepts and, therefore, overgeneralizes. The results of this study support this view.

The only significant differences found in an analysis of the kinds of errors made were between Condition 1 (rational set of examples and non-examples) and the control group on undergeneralization (on immediate acquisition only), and between Condition 2 (rational set of examples) and the control group on undergeneralization (on retention only). In both cases the control group made significantly more undergeneralization errors.

Neither of the two remaining dependent variables used in Experiment I, recognition of definitions and knowledge of relationships among concepts, was found to discriminate among the experimental conditions. There are two obvious explanations for this, both of which may in part account for the results. First, the concepts themselves may have been so difficult that merely presenting examples and non-examples did not provide subjects with enough information to infer the concept definitions or cognize the relationships among the concepts. Second. the dependent measures used may not have been valid or reliable. No reliability checks were run on the tests, and the number of items used to measure these variables was relatively small (six items for recognition of definitions and nine items for knowledge of relationships among concepts).*

Experiment II was essentially a replication of Experiment I, but definitions of the concepts were added to the treatment conditions. The results as measured by the dependent variables of correct classification of instances and recognition of

*Reliability estimates for the dependent measures were subsequently calculated and are included for the reader's information as Appendix F.

definitions (on both immediate acquisition and retention) showed consistently that when definitions were given with instances of the concepts, each of the treatment conditions differed significantly from the control condition but did not differ among themselves. Thus the addition of concept definitions had a facilitative effect (as measured by these specific dependent variables), which is consistent with the findings of Frayer (1970) and Merrill and Tennyson (1971). Not only did it negate any potential effects due to the number of examples presented, but it also compensated for the lack of non-examples in two of the treatment conditions (Conditions 2 and 3), which was found to be a crucial factor in Experiment I. These results indicate that possibly just providing a concept definition alone would in some cases be as facilitative in promoting concept learning as providing a concept definition with examples and non-examples.

On the dependent variable of knowledge of relationships among concepts only the difference between subjects who read lessons containing two examples with definitions and the control group was found to be significant. This suggests that possibly subjects who read lessons containing the rational set of both examples and non-examples with definitions, or just the rational set of examples with definitions, had been exposed to so much information about each concept that when they were asked to integrate the concepts and discern the relationships among them they confused what they had learned.

No significant differences were found between the treatment conditions on either over- or undergeneralization, on both immediate acquisition or retention. As this was identical to the results of Experiment I, the lack of significance was probably not due to any equalizing effects brought about by the presence of concept definitions. Possibly dividing errors into overgeneralization and undergeneralization categories is not as discriminating an analytic technique as Markle and Tiemann have hypothesized, but more likely the lack of significant differences was due to the overall difficulty of the concepts studied, or weaknesses in the experimental manipulations or dependent measures.

Significant differences on over- and undergeneralization errors were found, however, between treatment conditions and the control. Subjects who read lessons containing the rational set of examples and non-examples with definitions and subjects who read lessons containing just two examples with definitions overgeneralized significantly less than subjects in the control condition on both immediate acquisition and retention, while subjects in each of the treatment conditions undergeneralized significantly less than control subjects (both on immediate acquisition and retention).

In Experiment III the effect of presenting the rational set of examples and non-examples was contrasted with the effect of presenting the rational set of examples and non-examples with concept definitions, and with concept definitions and emphasis of relevant attributes. Contrary to the results of Experiment I, subjects who read lessons containing the rational set of examples and non-examples did not differ from the control group on the dependent variable of correct classification of instances (on immediate acquisition or retention). Indeed, subjects in this condition only differed from the control group on the dependent variable of undergeneralization errors on immediate acquisition, with subjects in the control group making significantly more undergeneralization errors. Possibly the effect of providing the rational set of examples and non-examples alone was not consistently found to be effective in promoting concept learning in Experiments I and III because the concepts were difficult ones and, as mentioned earlier, examples of them do not have easily discernible relevant attributes.

The addition of concept definitions and definitions plus emphasis to the rational set of examples and non-examples proved to be a facilitative instructional technique. For the dependent variables of correct classification of instances and recognition of definitions, subjects who read lessons containing the rational set of examples and non-examples with definitions or with definitions plus emphasis performed significantly better than subjects reading lessons containing only the rational set of examples and non-examples (on immediate acquisition), and better than the control subjects (on both immediate acquisition and retention for correct classification of instances and on immediate acquisition for recognition of definitions).

Interestingly, the addition of emphasis of relevant attributes to the rational set of examples and non-examples with concept definitions was only in one case found to be more facilitative than the rational set of examples and non-examples with definitions alone. Subjects who read lessons with emphasis differed significantly in performance from control subjects on the dependent variable of recognition of definitions on retention while none of the other treatment conditions did. Thus emphasis of relevant attributes did not have the significant effect which was hypothesized, indicating that the addition of definitions to the rational set was actually the critical factor. It should be pointed out, however, that the concepts of bilateral symmetry and rotational symmetry were defined in terms of only one relevant attribute, and translational symmetry was defined in terms of only two relevant attributes. It may very well be that emphasizing relevant attributes is only a facilitative technique when the concepts involved are more complex than those studied in the present series of experiments, and are defined in terms of several relevant attributes.

An analysis of the pattern of errors made in Experiment III showed that subjects who read lessons containing the rational set of examples and non-examples (Condition 1) both over- and undergeneralized more than subjects who read the same lessons but with concept definitions (Condition 2) on immediate acquisition, and they also overgeneralized more than subjects who read lessons containing the rational set of examples and non-examples, definitions and emphasis (Condition 3) on immediate acquisition. Again, this points out that the rational set of examples and non-examples alone was not as facilitative an instructional variable in Experiment III as the rational set of examples and non-examples with definitions, or with definitions and emphasis. Furthermore, while subjects in each of treatment Conditions 1, 2, and 3 undergeneralized significantly less than control subjects on immediate acquisition, only subjects in Conditions 2 and 3 undergeneralized significantly less than control subjects on retention. Additionally, on overgeneralization only subjects in Conditions 2 and 3 made significantly fewer errors than control subjects on immediate acquisition, although on retention there were no significant differences.

The fact that providing concept definitions or definitions plus emphasis with the rational set of examples and non-examples was consistently found in Experiment III to be more facilitative in promoting concept learning than the rational set of examples and non-examples alone is an important finding. It points out that Markle and Tiemann's approach to teaching concepts may not be the most effective method, at least for concepts similar to those studied in the present series of experiments.

In summary, the major findings of Experiments I-III can be broadly stated as follows. First, the use of non-examples which focus the subjects' attention on relevant attributes of the concept was found to be a facilitative instructional technique. Second, providing a concept definition generally compensated for presenting only the rational set of examples (with no non-examples), and negated any possible effects due to the number of examples given. Third, presenting the rational set of examples and non-examples was not consistently found to promote concept learning, but presenting the rational set with a concept definition generally was found to be effective. Finally, adding emphasis of relevant attributes to the rational set of examples and non-examples plus a concept definition was generally not found to significantly increase performance.

It is necessary to briefly point out, however, two basic limitations in the series of experiments which may have influenced the results and certainly limit their generalizability. As mentioned before, the concepts themselves were both difficult and defined only in terms of one or two relevant attributes. If other types of concepts had been studied the results might have been quite different. Additionally, the lessons and dependent measures had not been validated as to their effectiveness in either teaching or measuring concept learning. Indeed, the significant effects due to stratification level found on each experiment show that the effectiveness of the lessons for the individual subject was directly related to his ability to read, and the dependent variable of knowledge of relationships among concepts was rarely found to discriminate among the conditions. The results of Experiments I-III, therfore, may to some extent be a function of the experimental materials. However, the relative consistency of the results across experiments argues that the results are reliable, at least for the particular concepts studied.

Many questions dealing with the effects of the three instructional variables focused on in the present series of experiments on concept learning at the formal level still remain to be answered. Perhaps the most important of these is whether the same results as found here would be produced if different concepts were studied. Additionally, the use of other instructional variables in conjunction with examples and non-examples, definitions and emphasis should be investigated in an effort to determine what the optimal combination of instructional variables is in promoting concept learning, or whether such an optimal combination is in reality dependent upon the individual student or the concepts being studied.

REFERENCES

Amster, H., & Marascuilo, L. Effect of type of pretraining and variety of instances on children's concept learning. Journal of Experimental Child Psychology, 1965, 2, 192-204.

Anderson, R. C., & Kulhavy, R. W. Learning concepts from definitions. American Educational Research Journal, 1972, in press.

Bourne, L. E. Human conceptual behavior. Boston: Allyn and Bacon, 1966.

Clark, D. C. Teaching concepts in the classroom: A set of teaching prescriptions derived from experimental research. Journal of Educational Psychology, 1971, 62, 253-278.

Frayer, D. A. Effects of number of instances and emphasis of relevant attribute values on mastery of geometric concepts by fourth- and sixth-grade children. Technical Report Number 116. Wisconsin Research and Development Center for Cognitive Learning, The University of Wisconsin, Madison, 1970.

Gelfand, S. Effects of prior associations and task complexity upon the identification of concepts. Psychological Reports, 1958, 4, 567-574.

Hovland, C. I. A "communication analysis" of concept learning. Psychological Review, 1952, 59, 461-472.

Hovland, C. I., & Weiss, W. Transmission of information concerning concepts through positive and negative instances. Journal of Experimental Psychology, 1953, 45, 175-182.

Huttenlocker, J. Some effects of negative instances on the formation of simple concepts. Psychological Reports, 1962, 11, 35-42.

Klausmeier, H. J. Cognitive operations in concept learning. Paper presented at the meeting of the American Psychological Association, Washington, D.C., September, 1971.

Klausmeier, H. J., Davis, J. K., Ramsay, J. G., Fredrick, W. C., & Davies, M. H. Concept learning and problem solving: A bibliography, 1950-1964. Technical Report Number 1, Wisconsin Research and Development Center for Cognitive Learning, The University of Wisconsin, Madison, 1965.

Markle, S. M., & Tiemann, P. W. Really understanding concepts. Champaign, Illinois: Stipes, 1969.

Merrill, M. D. Necessary psychological conditions for defining instructional outcomes. In M. D. Merrill (Ed.), Instructional design readings. Englewood Cliffs, N. J.: Prentice Hall, 1971.

Merrill, M. D., & Tennyson, R. D. Concept acquisition as a function of relationships between positive and negative instances. Final Report of USOE Small Contract O-H-014, September, 1971.

Meyers, J. L. Fundamentals of experimental design. Boston: Allyn and Bacon, 1966.

Podell, H. A. Two processes of concept formation. Psychological Monographs, 1958, 72.

Remstad, R. C. Optimizing the response to a concept attainment test through sequential classroom experimentation. Technical Report Number 85. Wisconsin Research and Development Center for Cognitive Learning, The University of Wisconsin, Madison, 1969.

Smoke, K. L. Negative instances in concept learning. Journal of Experimental Psychology, 1933, 16, 583-588.

Smuckler, N. S. Concept formation as a function of method of presentation and ratio of positive to negative instances. Technical Report Number 27. Wisconsin Research and Development Center for Cognitive Learning, The University of Wisconsin, Madison, 1967.

Stern, C. Labeling and variety in concept identification with young children. Journal of Educational Psychology, 1965, 56, 235-240.

Tennyson, R. D. Correct classification, overgeneralization, undergeneralization, and misconception as a function of probability, divergency, and matching of instances in concept acquisition. Working Paper Number 19. Division of Communication Services, Brigham Young University, Provo, Utah, 1971.

Tennyson, R. D., Woolley, F. R., & Merrill, M. D. Exemplar and nonexemplar variables which produce correct classification behavior and specified classification errors. Journal of Educational Psychology, 1972, 63, 144-152.

Wittrock, M. C., Keislar, E. R., & Stern, C. Verbal cues in concept identification. Journal of Educational Psychology, 1964, 55, 195-200.

Appendices A and B have been omitted from this publication, but are available on microfilm from Memorial Library, University of Wisconsin, Madison, Wisconsin.

Appendix C

FIRST SESSION INSTRUCTIONS TO SUBJECTS

Instructions to Students

Good morning (afternoon).

My name is ____________________, and this is ____________________. We are working with some people at the University of Wisconsin in Madison who are very interested in finding better ways to help children learn about math. Today you will be able to help us by reading three math lessons and taking four tests. There is no reason to be worried about what you will be doing. You will not be graded on the tests. The information is just for us. But please try to do your best job.

In the brown envelopes which have been passed out to you are the lessons you will be reading. There are many different kinds of lessons. It may even seem that you are reading different lessons from everyone else. But don't worry about it because this is the way it is supposed to be.

Everyone's lessons are labeled Lesson I, Lesson II, and Lesson III. Please open your envelopes now and take out Lesson I. Only take out Lesson I and do not open it until I tell you to. (Wait until everyone has done this.) Please fill out the cover of Lesson I. Please print your name, the name of your school and teacher, your grade and today's date ________________ (Wait until everyone has finished).

Now turn to the first page where it says WORD LIST. These are some of the words which you may find in the lessons. Because they are a little unusual I would like to go over them with you. For instance, does anyone see the word ________? Good! What number is it? O.K., now let's all say it together. Fine! Now what about the word ________? (Continue until all the words have been pronounced and their numbers indicated.)

There may be other words in the lessons which are new to you. If you are having any trouble with a word just raise your hand and one of us will help you. Also, if you are asked any questions in the lessons you are reading, answer them right in the lesson booklet.

When you have finished reading Lesson I raise your hand and we will come around and collect your lesson and give you Test I. When you have finished with Test I turn it over and push it out of your way and then wait quietly until everyone has finished.

Does anyone have any questions?

Please work individually. Do <u>not</u> talk to one another. And please do the best job you can.

O.K., turn the page and begin reading.

Appendix D

SECOND SESSION INSTRUCTIONS TO SUBJECTS

Instructions to Students

Good morning (afternoon).

My name is ____________________ I am from the University of Wisconsin in Madison.

I am going to pass out some tests to you. They are like the tests you took two weeks ago on symmetry. Just as you weren't graded on those tests, you won't be graded on this test. The information is just for me and the people I work with in Madison. But please try to do your best job.

(Pass out the tests. Make sure that the children do not open the test booklets until told to do so.)

Now please fill in the cover of your test booklet. Write your name, teacher, school, grade and today's date --
(Write the date on the blackboard. Wait until everyone has finished before going on.)

Now open your booklets to the first page where it says general instructions. Please follow in your booklet while I read the instructions aloud. (Read the instructions.)

Please work independently on your test. Don't talk to one another. When you have finished your test, turn it over and wait quietly until everyone has finished.

Are there any questions? O.K., turn the page and begin. And do your best job.

(Collect the tests when everyone has finished. Then thank the kids and the teacher and leave.)

Appendix E

MEANS AND STANDARD DEVIATIONS FOR EACH PART

OF TESTS I-V BY CONDITION AND STRATIFICATION LEVEL

RAW SCORES

Experiments I-III

Means and Standard Deviations for Individual Parts

of Tests I-IV by Condition

Experiment I

Tests		Conditions 1 M	SD	2 M	SD	3 M	SD	4 M	SD
Test I									
Part I:	correct	12.4500	2.2907	11.0000	2.0000	12.2500	2.7363	11.5000	2.6739
	under	3.2000	1.9131	3.5263	1.3126	3.3000	1.6763	3.8500	1.8241
	over	4.3500	1.5898	4.8421	1.4961	4.4000	1.6852	4.5500	1.4992
Part II:	correct	0.5500	0.4975	0.4737	0.4993	0.4500	0.4975	0.4500	0.4975
Test II									
Part I:	correct	11.2000	2.7677	11.1053	1.9707	11.0000	3.9243	10.0000	2.4900
	under	3.8500	2.1042	3.5263	1.6016	4.6500	2.4955	5.1500	1.9046
	over	4.9500	1.7741	5.3158	1.3785	4.2500	1.9462	4.8500	1.4239
Part II:	correct	0.1500	0.3571	0.1053	0.3069	0.1500	0.3571	0.2000	0.4000
Test III									
Part I:	correct	13.9500	2.7290	12.0526	3.2682	12.1500	2.8509	11.4000	2.7641
	under	2.1000	1.5133	2.9474	2.0384	3.6500	2.0069	3.5000	2.1095
	over	3.9000	1.8947	5.0000	1.8064	4.2000	1.6912	5.0500	1.6271
Part II:	correct	0.2500	0.4330	0.1579	0.3646	0.1500	0.3571	0.1500	0.3571
Test IV									
Part I:	correct	2.4500	1.2440	2.1579	1.0394	2.1000	1.4107	1.7000	0.9000
Part II:	correct	0.8500	0.9631	0.8947	0.6406	0.6500	1.0137	0.6000	0.6633

Means and Standard Deviations for Individual Parts

of Test V by Condition

Experiment I

Tests	Conditions 1 M	SD	2 M	SD	3 M	SD	4 M	SD
Test V - Part I								
A: correct	12.5000	2.4393	12.3684	2.9773	12.7000	3.1000	11.4500	3.1855
under	3.8000	1.9900	3.6842	2.1039	3.2000	1.7493	4.5000	1.8574
over	3.7000	1.4526	3.6316	1.9523	3.9000	1.9209	4.0000	1.8166
B: correct	11.2500	2.8085	13.0000	2.3170	11.9000	3.5763	10.7500	2.0463
under	3.7500	2.1418	2.8421	1.6941	3.8000	2.1354	4.3000	1.4526
over	4.7000	1.8466	4.1579	1.5648	4.0000	1.5166	4.9000	1.4457
C: correct	12.7500	2.9304	12.4211	3.1841	11.5500	3.1060	11.1000	2.3431
under	2.9500	1.9868	2.7895	1.7941	3.7500	2.2555	4.2000	2.1354
over	4.0000	1.9748	4.7895	2.0921	4.3000	1.7349	4.6500	1.3143
Part II								
correct	0.5500	0.8047	1.1053	0.9676	0.5000	0.8062	0.7500	0.6225

Means and Standard Deviations for Individual Parts

of Test I-IV by Stratification Level

Experiment I

Tests	Stratification Levels					
	1		2		3	
	M	SD	M	SD	M	SD
Test I						
Part A: correct	10.7407	2.2376	11.5926	1.8710	13.2000	2.7568
: under	3.7407	1.6687	3.9259	1.4638	2.6800	1.7600
: over	5.2593	1.2047	4.2593	1.5538	4.0400	1.6848
Part B: correct	0.4074	0.4914	0.4074	0.4914	0.6400	0.4800
Test II						
Part A: correct	10.7037	2.1225	10.0000	2.4944	11.8400	3.7060
: under	4.4444	1.8725	4.7037	2.0515	3.7200	2.4087
: over	4.8519	1.1771	5.2593	1.6687	4.3600	2.0373
Part B: correct	0.1481	0.3552	0.1481	0.3552	0.1600	0.3666
Test III						
Part A: correct	11.3333	2.6667	11.5556	2.4545	14.4400	3.0342
: under	3.3704	2.0394	3.5556	1.7069	2.1600	2.0333
: over	5.2593	1.7970	4.8889	1.3966	3.3600	1.6942
Part B: correct	0.1852	0.3884	0.1111	0.3143	0.2400	0.4271
Test IV						
Part A: correct	1.8148	1.0554	1.8519	1.0436	2.6800	1.2875
Part B: correct	0.3704	0.6175	0.5556	0.5666	1.3600	0.9749

Means and Standard Deviations for Individual Parts

of Test V by Stratification Level

Experiment I

Tests		Stratification Levels					
		1		2		3	
		M	SD	M	SD	M	SD
Test V							
Part A:	correct	11.6667	2.0728	11.0370	2.8478	14.2000	2.9799
:	under	3.7778	1.5713	4.5926	1.9486	2.9600	2.0684
:	over	4.2222	1.8325	4.3333	1.5396	2.8000	1.6000
Part B:	correct	11.1852	2.4950	10.7037	2.7864	13.3600	2.6364
:	under	3.6296	1.5904	4.6296	1.9273	2.7200	1.8443
:	over	4.7407	1.6908	4.6296	1.6136	3.9200	1.4945
Part C:	correct	11.3704	2.9332	11.2222	2.1315	13.3600	3.3091
:	under	3.5556	2.2662	4.0370	1.5026	2.6400	2.3131
:	over	4.7037	1.9209	4.5926	1.8907	3.9600	1.5357
Part D:	correct	0.3704	0.4829	0.8148	0.8623	1.0000	0.9798

Means and Standard Deviations for Individual Parts

of Tests I-IV by Condition

Experiment II

Tests	Conditions 1 M	1 SD	2 M	2 SD	3 M	3 SD	4 M	4 SD
Test I								
Part I: correct	13.6071	2.9682	13.1538	3.4718	15.1786	3.0944	11.1379	2.6356
: under	2.4286	2.0429	2.4615	1.6227	1.7500	1.7652	3.8621	1.9779
: over	3.7857	2.0417	4.3077	2.4459	3.0714	1.7714	4.9310	1.7991
Part II: correct	0.7500	0.4330	0.7692	0.4213	0.7857	0.4103	0.4828	0.4997
Test II								
Part I: correct	13.9286	2.4774	14.8462	2.6267	14.1429	3.0789	11.8966	2.9751
: under	2.6786	1.5131	1.4231	1.2457	2.8571	2.2315	3.6897	1.7440
: over	3.2500	1.7449	3.7308	1.8305	2.8929	1.7390	4.3448	2.0890
Part II: correct	0.6071	0.4884	0.6923	0.4615	0.8571	0.3499	0.2069	0.4051
Test III								
Part I: correct	13.6071	2.5543	14.6154	2.7467	13.5357	2.8596	11.5172	2.0615
: under	2.7143	2.1020	1.5385	1.3368	2.2500	1.6610	3.8276	1.8582
: over	3.5357	1.6362	3.8462	1.9941	4.2143	1.8776	4.6552	1.2939
Part II: correct	0.7143	0.4518	0.5385	0.4985	0.6786	0.4670	0.3793	0.4852
Test IV								
Part I: correct	2.7143	2.1020	2.6154	1.6190	3.0000	1.8323	1.7241	1.0795
Part II: correct	1.6071	1.1753	1.5385	1.1174	2.0714	1.0996	1.0000	1.1142

Means and Standard Deviations for Individual Parts

of Test V by Condition

Experiment II

Test	Conditions							
	1		2		3		4	
	M	SD	M	SD	M	SD	M	SD
Test V - Part I								
A: correct	14.4643	2.8219	12.8462	3.6658	14.1071	3.6775	10.5862	2.2670
: under	2.5714	1.9898	3.3846	2.3383	2.6071	2.2731	4.7931	1.5840
: over	2.9643	2.0438	3.6538	2.2861	3.2500	2.4440	4.6207	2.0071
B: correct	13.6429	2.9058	13.9231	3.4743	13.7500	2.8862	10.6552	3.5552
: under	2.8214	2.2209	2.5385	2.1703	3.0000	2.1044	4.4483	2.1429
: over	3.5357	2.0438	3.5385	2.1345	3.2500	1.7652	4.8621	2.1928
C: correct	14.0000	2.2361	13.8077	2.8014	13.5714	2.9085	11.6552	2.7073
: under	2.3929	1.6331	2.1154	1.6011	2.6071	1.9150	3.2414	1.7151
: over	3.6071	1.6975	4.0385	1.8288	3.8214	1.7332	5.0000	1.9119
Part II								
correct	1.6429	0.9340	1.5385	1.1174	1.7500	1.1220	0.8276	0.9850

Means and Standard Deviations for Individual Parts of Test I-IV by Stratification Level Experiment II

Tests		Stratification Levels 1 M	1 SD	2 M	2 SD	3 M	3 SD
Test I							
Part A:	correct	11.3056	2.6752	14.1842	2.9545	14.1892	3.5855
:	under	3.4444	2.2785	2.0789	1.5957	2.4324	1.8965
:	over	5.2222	1.8873	3.6842	1.9481	3.2162	2.0419
Part B:	correct	0.5833	0.4930	0.7105	0.4535	0.7838	0.4117
Test II							
Part A:	correct	12.7778	2.5068	13.4211	2.8063	14.7838	3.3137
:	under	2.9167	1.8615	2.8421	1.7551	2.3243	2.0409
:	over	4.2778	1.7890	3.6579	1.9368	2.7568	1.7768
Part B:	correct	0.5000	0.5000	0.5526	0.4972	0.7027	0.4571
Test III							
Part A:	correct	11.8889	2.4920	13.6842	2.6168	14.2162	2.7522
:	under	3.3333	2.2485	2.4737	1.5172	2.0541	1.8299
:	over	4.7778	1.6349	3.8421	1.7400	3.6216	1.6982
Part B:	correct	0.4167	0.4930	0.5526	0.4972	0.7568	0.4290
Test IV							
Part A:	correct	1.3889	1.0872	2.4211	1.2697	3.6757	1.9872
Part B:	correct	0.6389	0.7871	1.7105	1.0236	2.2703	1.1065

Means and Standard Deviations for Individual Parts

of Test V by Stratification Level

Experiment II

Tests		Stratification Levels					
		1		2		3	
		M	SD	M	SD	M	SD
Test V							
Part A:	correct	11.3056	2.7970	13.5263	3.4848	14.0541	3.5485
:	under	4.2778	2.2062	2.7105	1.7458	3.1081	2.4581
:	over	4.3333	2.2608	3.7368	2.4675	2.8378	1.8382
Part B:	correct	12.0556	3.1265	12.5000	3.2586	14.2973	3.6751
:	under	3.4167	1.8008	3.4474	2.3697	2.8108	2.5451
:	over	4.5000	2.0616	4.0526	2.1879	2.8919	1.8126
Part C:	correct	12.0556	2.7983	13.2105	2.5043	14.4054	2.7159
:	under	3.0000	1.6667	2.7368	1.6965	2.0811	1.8215
:	over	4.8889	1.7916	4.0000	1.8496	3.5135	1.7183
Part D:	correct	0.8333	0.9280	1.4737	1.0192	1.9730	1.0523

Means and Standard Deviations for Individual Parts

of Tests I-IV by Condition

Experiment III

Tests	Conditions 1 M	SD	2 M	SD	3 M	SD	4 M	SD
Test I								
Part I: correct	11.7083	3.2078	15.5714	3.0170	16.6250	2.7358	11.2917	3.3600
: under	3.5000	1.7559	1.6429	1.5169	1.2083	1.2576	4.0417	2.0912
: over	4.7917	1.9786	2.7857	2.0763	2.1667	1.7480	4.6667	2.0138
Part II: correct	0.5000	0.5000	0.7143	0.4518	0.7500	0.4330	0.5417	0.4983
Test II								
Part I: correct	12.6667	3.5198	14.1786	3.0479	14.5000	3.2660	12.1250	3.9509
: under	2.8333	1.9076	2.0714	1.7914	2.6667	1.9293	4.1667	2.4944
: over	4.4167	2.0599	3.7500	1.8637	2.8333	2.1538	3.6667	1.9720
Part II: correct	0.1667	0.3727	0.6429	0.4792	0.7500	0.4330	0.2917	0.4545
Test III								
Part I: correct	14.0833	2.9849	14.6071	2.6905	14.0000	2.5000	11.0417	2.8647
: under	2.7083	1.8815	2.4643	1.6579	2.4167	1.6051	4.1250	1.9432
: over	3.2083	1.5270	2.9286	1.5336	3.5417	2.1598	4.7917	1.7315
Part II: correct	0.3750	0.4841	0.6786	0.4670	0.6667	0.4714	0.2917	0.4545
Test IV								
Part I: correct	2.4583	1.4994	3.2143	1.9704	3.5833	1.9983	2.2917	1.8366
Part II: correct	1.0000	1.0000	1.8929	1.1129	2.5000	0.9129	1.1250	1.0533

Means and Standard Deviations for Individual Parts

of Test V by Condition

Experiment III

Tests	Conditions							
	1		2		3		4	
	M	SD	M	SD	M	SD	M	SD
Test V - Part I								
A: correct	11.7917	3.2657	14.8571	3.4715	16.1250	2.8035	11.3333	3.2998
: under	3.9583	2.1111	2.2143	2.1606	1.5417	1.6325	4.3333	2.1922
: over	4.2500	1.7854	2.8929	1.9882	2.3333	1.5986	4.3333	1.8409
B: correct	12.6250	3.7060	14.3214	2.9887	13.5417	3.7414	13.3333	3.4841
: under	3.1250	2.3684	1.9286	1.9987	3.1250	2.4206	3.3750	2.0578
: over	4.2500	1.8540	3.7500	1.8444	3.3333	2.3570	3.0417	2.0100
C: correct	13.8333	3.0231	14.1786	2.6465	13.7917	3.1222	11.8750	3.0864
: under	2.3750	2.0169	2.3214	1.7124	2.5833	2.0599	3.3333	1.8409
: over	3.7500	1.5612	3.4643	1.8416	3.4583	1.8253	4.7917	2.0203
Part II								
correct	1.2917	1.2741	2.0000	1.0351	2.2083	1.0793	1.2500	1.1273

Means and Standard Deviations for Individual Parts

of Test I-IV by Stratification Level

Experiment III

Tests	Stratification Levels					
	1		2		3	
	M	SD	M	SD	M	SD
Test I						
Part A: correct	13.0286	3.7606	13.2000	3.4775	15.2857	3.8587
: under	3.0857	1.9766	2.7000	1.7916	1.9143	2.1695
: over	3.8857	2.1748	4.1000	2.2561	2.8000	2.1620
Part B: correct	0.4286	0.4949	0.5667	0.4955	0.8857	0.3182
Test II						
Part A: correct	11.6571	2.8779	13.5333	3.3935	15.0286	3.5896
: under	3.5143	1.9621	2.9333	2.1899	2.2571	2.1954
: over	4.7714	1.7902	3.5333	1.7839	2.6857	2.0670
Part B: correct	0.4000	0.4899	0.4667	0.4989	0.5429	0.4982
Test III						
Part A: correct	12.1429	2.6741	14.0000	3.0441	14.3714	3.0714
: under	3.6000	1.4182	2.9333	1.9989	2.2000	1.9828
: over	4.2571	1.9025	3.0667	1.5041	3.3714	1.9725
Part B: correct	0.4571	0.4982	0.5000	0.5000	0.5714	0.4949
Test IV						
Part A: correct	2.2000	1.4890	2.6000	1.5406	3.8571	2.1797
Part B: correct	1.2286	1.0443	1.4667	1.2311	2.2000	1.0637

Means and Standard Deviations for Individual Parts
of Test V by Stratification Level
Experiment III

Tests	Stratification Levels					
	1		2		3	
	M	SD	M	SD	M	SD
Test V						
Part A: correct	12.3143	3.6938	13.6333	3.0603	14.8000	4.0553
: under	3.6000	2.2194	3.1333	1.9448	2.2286	2.5644
: over	4.0857	2.0890	3.2333	1.5206	2.9429	2.1104
Part B: correct	11.7714	3.2784	13.3000	2.9905	15.3714	3.2609
: under	3.6571	2.3292	2.9000	1.8682	2.0000	2.2678
: over	4.5714	1.9315	3.6000	1.9933	2.6286	1.7902
Part C: correct	12.0286	2.8131	13.7000	3.0238	14.6571	2.8480
: under	3.2857	1.6137	2.5333	2.0287	2.0857	1.9910
: over	4.5429	1.8415	3.7667	1.9947	3.2286	1.6228
Part D: correct	1.4286	1.1029	1.5333	1.1175	2.1143	1.2597

Appendix F

HOYT RELIABILITY ESTIMATES FOR

DEPENDENT MEASURES

Experiments I-III

Hoyt Reliability Estimates for

Dependent Measures

Experiments I-III

Dependent Measures	Experiments		
	I	II	III
Part I, Tests I-III (correct classification)	.65	.78	.84
Part II, Tests I-IV (recog. of definitions)	.35	.74	.79
Part I, Test IV (knowl. of relationships)	.08	.51	.59
Part I, Test V (sections A-C) (correct classification)	.66	.82	.85
Part II, Test V (recog. of definitions)	.37	.58	.74

National Evaluation Committee

Helen Bain
Immediate Past President
National Education Association

Lyle E. Bourne, Jr.
Institute for the Study of Intellectual Behavior
Universiy of Colorado

Jeanne S. Chall
Graduate School of Education
Harvard University

Francis S. Chase
Department of Education
University of Chicago

George E. Dickson
College of Education
University of Toledo

Hugh J. Scott
Superintendent of Public Schools
District of Columbia

H. Craig Sipe
Department of Instruction
State University of New York

G. Wesley Sowards
Dean of Education
Florida International University

Benton J. Underwood
Department of Psychology
Northwestern University

Robert J. Wisner
Mathematics Department
New Mexico State University

Executive Committee

William R. Bush
Director of Program Planning and Management
and Deputy Director, R & D Center

Herbert J. Klausmeier, Committee Chairman
Director, R & D Center

Wayne Otto
Principal Investigator
R & D Center

Robert G. Petzold
Professor of Music
University of Wisconsin

Richard A. Rossmiller
Professor of Educational Administration
University of Wisconsin

James E. Walter
Coordinator of Program Planning
R & D Center

Russell S. Way, ex officio
Program Administrator, Title III ESEA
Wisconsin Department of Public Instruction

Faculty of Principal Investigators

Vernon L. Allen
Professor of Psychology

Frank H. Farley
Associate Professor
Educational Psychology

Marvin J. Fruth
Associate Professor
Educational Administration

John G. Harvey
Associate Professor
Mathematics

Frank H. Hooper
Associate Professor
Child Development

Herbert J. Klausmeier
Center Director
V. A. C. Henmon Professor
Educational Psychology

Stephen J. Knezevich
Professor
Educational Administration

Joel R. Levin
Associate Professor
Educational Psychology

L. Joseph Lins
Professor
Institutional Studies

Wayne Otto
Professor
Curriculum and Instruction

Thomas A. Romberg
Associate Professor
Curriculum and Instruction

Peter A. Schreiber
Assistant Professor
English

Richard L. Venezky
Associate Professor
Computer Science

Alan M. Voelker
Assistant Professor
Curriculum and Instruction

Larry M. Wilder
Assistant Professor
Communication Arts